BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:57:01 ; Search time 8.18182 Seconds
(without alignments)
528.928 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	103.5	48.6	210	2	A25550	histone H1 - sea u
2	98	46.0	394	2	F90725	membrane spanning
3	98	46.0	394	2	G85576	membrane spanning
4	98	46.0	421	2	JV0057	tolA protein - Esc
5	94.5	44.4	239	2	AE1317	hypothetical prote
6	90	42.3	243	2	AE1689	hypothetical prote
7	90	42.3	1390	2	S51364	sperm tail-specifi
8	89.5	42.0	211	2	A28100	histone H1-beta, e
9	89.5	42.0	214	2	G70673	probable hupB - My
10	88	41.3	899	2	B87553	DNA topoisomerase
11	86.5	40.6	384	2	B43592	outer membrane pro
12	86	40.4	199	2	F70742	hypothetical prote
13	86	40.4	344	2	S34153	mst101-1 protein -
14	86	40.4	347	2	E83525	TolA protein PA097
15	85	39.9	182	2	S61926	histone H1 homolog
16	82	38.5	248	1	HSUR1P	histone H1, gonada
17	82	38.5	262	2	T35503	1-acylglycerol-3-p
18	82	38.5	568	2	T39675	asparaginyl-trna s
19	81.5	38.3	311	2	T17698	hypothetical prote
20	80.5	37.8	277	2	T34625	probable NLP/P60 f
21	80.5	37.8	445	2	T50972	probable zuotin [i
22	80.5	37.8	952	2	T36664	probable DNA topoi
23	80	37.6	194	1	HSTR1	histone H1 - trout
24	80	37.6	206	1	HSTR1R	histone H1 - rainb
25	80	37.6	291	2	AB1487	hypothetical prote
26	79.5	37.3	217	2	A26721	histone H1-gamma,
27	79	37.1	62	2	T30977	hypothetical prote
28	79	37.1	205	2	S19114	cgcr-1 protein - C
29	79	37.1	244	2	S40436	histone H1 - midge
30	79	37.1	347	2	A81794	probable rotamase
31	78	36.6	217	2	S29309	hypothetical prote
32	78	36.6	231	2	S59589	histone H1 - Chlam
33	78	36.6	287	2	S45662	histone H1 - tomat
34	78	36.6	309	2	G83013	polyhydroxyalkanoa
35	77.5	36.4	376	2	AG0592	tolA protein [impo
36	77	36.2	220	2	A28456	histone H1.10 - ch
37	77	36.2	226	1	S51660	histone H1-5 [vali
38	77	36.2	523	2	B38145	invariant surface
39	77	36.2	580	2	B70868	probable transfera
40	76	35.7	229	2	I51227	histone H1A - Afri
41	76	35.7	241	2	JN0748	histone H1-II - Vo
42	76	35.7	388	2	AC0138	TolA colicin impor
43	75.5	35.4	220	2	JC5954	ribosomal protein
44	75.5	35.4	924	2	T06636	hypothetical prote
45	75	35.2	212	2	A28470	histone H1 - mouse

ALIGNMENTS

RESULT 1
A25550
histone H1 - sea urchin (Lytechinus pictus)
C;Species: Lytechinus pictus (painted urchin)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
C;Accession: A25550
R;Knowles, J.A.; Childs, G.J.
Nucleic Acids Res. 14, 8121-8133, 1986
A;Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus an
A;Reference number: A25550; MUID:87040778; PMID:3022245
A;Accession: A25550
A;Molecule type: DNA
A;Residues: 1-210 <KNO>
A;Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617
C;Superfamily: histone H1
C;Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 48.6%; Score 103.5; DB 2; Length 210;
Best Local Similarity 65.1%; Pred. No. 0.0053;
Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

```
Qy          1 AKKYAKKAKAEK-AKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              ||| |||  |:| |||| |   ||||||  | ||:| |||:||
Db        166 AKKAAKKPAAKKPAKKAAKKPAAKKAAKPAKKAAKKPAAKKAA 208
```

RESULT 2
F90725
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90725
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90725
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34197.1; PID:g13360233; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0774

Query Match 46.0%; Score 98; DB 2; Length 394;
Best Local Similarity 65.3%; Pred. No. 0.028;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

```
Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || |:|| ||||   ||  | |||  ||||   ||||||||| :|| ||
```

```
Db        220 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 267


RESULT 3
G85576
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C;Accession: G85576
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <STO>
A;Cross-references: GB:AE005174; NID:g12513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:Z09
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: tolA

  Query Match             46.0%;  Score 98;  DB 2;  Length 394;
  Best Local Similarity   65.3%;  Pred. No. 0.028;
  Matches   32;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || |:|| ||||   ||  | |||  ||||   ||||||||| :|| ||
Db        220 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 267


RESULT 4
JV0057
tolA protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002
C;Accession: JV0057; B64810
R;Levengood, S.K.; Webster, R.E.
J. Bacteriol. 171, 6600-6609, 1989
A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their produ
A;Reference number: JV0057; MUID:90078104; PMID:2687247
A;Accession: JV0057
A;Molecule type: DNA
A;Residues: 1-421 <LEV>
A;Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019
A;Experimental source: strain JM105
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64810
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-421 <BLAT>
A;Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73833.1; PID:g1786960;
A;Experimental source: strain K-12, substrain MG1655
C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach t
C;Genetics:
A;Gene: tolA
A;Map position: 17 min
A;Start codon: GTG
C;Keywords: nucleotide binding; P-loop; transmembrane protein
F;14-34/Domain: transmembrane #status predicted <MSS>
F;78-301/Domain: helical #status predicted <HSR>
F;355-362/Region: nucleotide-binding motif A (P-loop)

  Query Match             46.0%;  Score 98;  DB 2;  Length 421;
  Best Local Similarity   65.3%;  Pred. No. 0.029;
  Matches   32;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || |:|| ||||   ||  | |||  ||||   ||||||||| :|| ||
Db        247 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 294


RESULT 5
AE1317
hypothetical protein lmo1941 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1317
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1317
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00019.1; PID:g16411394; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1941

  Query Match             44.4%;  Score 94.5;  DB 2;  Length 239;
  Best Local Similarity   56.2%;  Pred. No. 0.04;
  Matches   27;  Conservative    7;  Mismatches    9;  Indels    5;  Gaps    2;

Qy          2 KKYAKKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAKEAAYE 44
              |  |:||:|:| |:   | ||| |||   |  |||||:||||::|| |
Db        124 KAAAEKAEADKKKQEEDAVKAANAKKEQEAAEEKAAADKAAAEKAAAE 171


RESULT 6
AE1689
hypothetical protein homolog lin2055 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1689
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1689
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97285.1; PID:g16414556; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2055

  Query Match             42.3%;  Score 90;  DB 2;  Length 243;
  Best Local Similarity   45.5%;  Pred. No. 0.11;
  Matches   30;  Conservative    3;  Mismatches   11;  Indels   22;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKA--------------------AKYEKAAAEKAAA 38
              ||| | :|| |||| :  |||| | |                    |  ||||||||||
Db        106 AKKAAEEKAAAEKAAEEKKAAEEKAAADKKSQEDEAAKAAAAKKEQEAAEEKAAAEKAAA 165

Qy         39 KEAAYE 44
                :|| |
Db        166 DKAAKE 171
```

```
RESULT 7
S51364
sperm tail-specific protein mst101(2) - fruit fly (Drosophila hydei)
C;Species: Drosophila hydei
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S51364; S34154
R;Neesen, J.; Padmanabhan, S.; Buenemann, H.
Eur. J. Biochem. 225, 1089-1095, 1994
A;Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represe
lpha-helical rods within the extremely elongated spermatozoa of Drosophila hydei.
A;Reference number: S51364; MUID:95045538; PMID:7957199
A;Accession: S51364
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1390 <NEE>
A;Cross-references: EMBL:X73481
R;Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A;Reference number: S34153
A;Accession: S34154
A;Molecule type: DNA
A;Residues: 1-163,'E',164-236,'Q',237-254,257-320,'E',321-1390 <NEW>
A;Cross-references: EMBL:X73481; NID:g313201; PID:g313202
C;Genetics:
A;Gene: mst101(2)
A;Cross-references: FlyBase:FBgn0011816

  Query Match             42.3%;  Score 90;  DB 2;  Length 1390;
  Best Local Similarity   57.8%;  Pred. No. 0.41;
  Matches   26;  Conservative    4;  Mismatches   11;  Indels    4;  Gaps    2;

Qy          2 KKYAKKAKAEK-AKKAYKAAEAKK---AAKYEKAAAEKAAAKEAA 42
              |: |:| | || |||  :||| ||   ||| || ||||   :|||
Db        662 KETAEKKKCEKAAKKRKEAAEKKKCAEAAKKEKEAAEKKKCEEAA 706


RESULT 8
A28100
histone H1-beta, embryonic - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 23-Feb-1997
C;Accession: A28100
R;Lai, Z.C.; Childs, G.
Mol. Cell. Biol. 8, 1842-1844, 1988
A;Title: Characterization of the structure and transcriptional patterns of the gene encc
A;Reference number: A28100; MUID:88246461; PMID:2837660
A;Accession: A28100
A;Molecule type: DNA
A;Residues: 1-211 <LAI>
A;Cross-references: GB:M20314
C;Superfamily: histone H1
C;Keywords: chromosomal protein; DNA binding; embryo; nucleosome; nucleus

  Query Match             42.0%;  Score 89.5;  DB 2;  Length 211;
  Best Local Similarity   61.9%;  Pred. No. 0.11;
  Matches   26;  Conservative    3;  Mismatches   10;  Indels    3;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              :||  || |   |||| | | ||||||  | ||:| |||:||
Db        139 SKKTTKKVKKPAAKKAKKPA-AKKAAK--KPAAKKPAAKKAA 177


RESULT 9
G70673
probable hupB - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70673
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70673
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-214 <COL>
A;Cross-references: GB:Z83018; GB:AL123456; NID:g3261671; PIDN:CAB05427.1; PID:g1694845
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: hupB
C;Superfamily: histone H1

  Query Match             42.0%;  Score 89.5;  DB 2;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.11;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  |  |   : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 10
B87553
DNA topoisomerase I [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87553
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87553
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-899 <STO>
A;Cross-references: GB:AE005673; NID:g13423998; PIDN:AAK24422.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2451
C;Superfamily: bacterial type I DNA topoisomerase

  Query Match             41.3%;  Score 88;  DB 2;  Length 899;
  Best Local Similarity   58.7%;  Pred. No. 0.45;
  Matches   27;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYAKKAKAEKAK-KAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| | |:| ||   | ||| |  :| ||:| |||:|| :|
Db        836 AKKPAKKAAATKSKAKAESDAPAKKTAA-KKPAAKKPAAKKAAPKA 880


RESULT 11
B43592
outer membrane protein TmpB - Treponema phagedenis
C;Species: Treponema phagedenis
C;Date: 30-Jan-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999
C;Accession: B43592
R;Yelton, D.B.; Limberger, R.J.; Curci, K.; Malinosky-Rummell, F.; Slivienski, L.; Schou
Infect. Immun. 59, 3685-3693, 1991
A;Title: Treponema phagedenis encodes and expresses homologs of the Treponema pallidum T
A;Reference number: A43592; MUID:91372983; PMID:1894368
A;Accession: B43592
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-384 <YEL>
A;Cross-references: GB:M58563; NID:g155066; PIDN:AAA27480.1; PID:g155067
A;Note: the authors translated the codon TTC for residue 316 as Tyr, and CGA for residue
C;Keywords: membrane protein

  Query Match             40.6%;  Score 86.5;  DB 2;  Length 384;
  Best Local Similarity   54.5%;  Pred. No. 0.32;
  Matches   24;  Conservative    6;  Mismatches   13;  Indels    1;  Gaps    1;
```

```
Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYE 44
              ||: | | || | | | : |  :|||| :||| |||| ::|| |
Db        178 AKEKAAKEKAAKDKAAKEKAAKEKAAK-DKAAKEKAAKEKAARE 220
```

RESULT 12
F70742
hypothetical protein Rv0475 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: F70742
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70742
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-199 <COL>
A;Cross-references: GB:Z77162; GB:AL123456; NID:g3261606; PIDN:CAB00936.1; PID:e255021;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0475

  Query Match   40.4%; Score 86; DB 2; Length 199;
  Best Local Similarity   65.6%; Pred. No. 0.21;
  Matches   21; Conservative   4; Mismatches   7; Indels   0; Gaps   0;

```
Qy         11 EKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              :||  | ||| |||||  :||||:|| ||:||
Db        161 KKAAPAKKAAPAKKAAPAKKAAAKKAPAKKAA 192
```

RESULT 13
S34153
mst101-1 protein - fruit fly (Drosophila hydei)
C;Species: Drosophila hydei
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C;Accession: S34153
R;Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A;Reference number: S34153
A;Accession: S34153
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-344 <NEE>
A;Cross-references: EMBL:X73480; NID:g313199; PID:g313200
C;Genetics:
A;Gene: FlyBase:Dhyd/mst101
A;Cross-references: FlyBase:FBgn0011816
C;Superfamily: neurofilament triplet H protein

  Query Match   40.4%; Score 86; DB 2; Length 344;
  Best Local Similarity   57.8%; Pred. No. 0.33;
  Matches   26; Conservative   3; Mismatches   12; Indels   4; Gaps   2;

```
Qy          2 KKYAKKAK-AEKAKKAYKAAEAKK---AAKYEKAAAEKAAAKEAA 42
              |: |:| | || |||  :||| ||   ||| || ||||    |||
Db         69 KEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAA 113
```

RESULT 14
E83525
TolA protein PA0971 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83525
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-347 <STO>
A;Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: tolA; PA0971

  Query Match   40.4%; Score 86; DB 2; Length 347;
  Best Local Similarity   51.1%; Pred. No. 0.33;
  Matches   24; Conservative   6; Mismatches   15; Indels   2; Gaps   1;

```
Qy          1 AKKYAKKAKAEKAKKA--YKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |:| |:  ||::||||   |||| || |   |  ||  | |:|| :|
Db        125 AQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKRAEDEAKKKAAEDA 171
```

RESULT 15
S61926
histone H1 homolog - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 23-Jul-1996 #sequence_revision 06-Sep-1996 #text_change 24-Nov-1999
C;Accession: S61926; S69327
R;Scarlato, V.; Arico, B.; Goyard, S.; Ricci, S.; Manetti, R.; Prugnola, A.; Manetti, R.
Mol. Microbiol. 15, 871-881, 1995
A;Title: A novel chromatin-forming histone H1 homologue is encoded by a dispensable and
A;Reference number: S61926; MUID:95319329; PMID:7596289
A;Accession: S61926
A;Molecule type: DNA
A;Residues: 1-182 <SCA1>
A;Cross-references: EMBL:L37438; NID:g777717; PIDN:AAB59120.1; PID:g777718
A;Accession: S69327
A;Molecule type: protein
A;Residues: 'X',3-39 <SCA2>
C;Genetics:
A;Gene: bpH1
C;Superfamily: histone H1

  Query Match   39.9%; Score 85; DB 2; Length 182;
  Best Local Similarity   61.4%; Pred. No. 0.25;
  Matches   27; Conservative   4; Mismatches   9; Indels   4; Gaps   3;

```
Qy          1 AKKYAKKAKAEK--AKKAY-KAAEAKKAAKYEKAAAEKAAAKEA 41
              ||| |||| |:|  ||||  | | ||||   :|| |:|| ||:|
Db         58 AKKVAKKAVAKKAVAKKAVAKKAVAKKAVA-KKAVAKKAVAKKA 100
```

Search completed: January 21, 2004, 09:01:39
Job time : 9.18182 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:56:26 ; Search time 6.13636 Seconds
(without alignments)
344.863 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	103.5	48.6	210	1	H1_LYTPI	P06144 lytechinus
2	98	46.0	421	1	TOLA_ECOLI	P19934 escherichia
3	90	42.3	1391	1	MST2_DROHY	Q08696 drosophila
4	89.5	42.0	208	1	DBH_MYCSM	Q9zhc5 mycobacteri
5	89.5	42.0	211	1	H1B_STRPU	P15869 strongyloce
6	89.5	42.0	214	1	DBH_MYCTU	P95109 mycobacteri
7	86.5	40.6	384	1	TMPB_TREPH	P29720 treponema p
8	86	40.4	168	1	RS16_COREF	Q8fp30 corynebacte
9	86	40.4	198	1	HBHA_MYCTU	Q11142 mycobacteri
10	86	40.4	344	1	MST1_DROHY	Q08695 drosophila
11	86	40.4	347	1	TOLA_PSEAE	P50600 pseudomonas
12	85.5	40.1	205	1	DBH_MYCBO	Q9xb18 mycobacteri
13	82	38.5	165	1	RS16_CORGL	Q8nnx3 corynebacte
14	82	38.5	248	1	H1_PARAN	P02256 parechinus
15	81	38.0	235	1	H1E_CHIPA	P40262 chironomus
16	80.5	37.8	952	1	TOP1_STRCO	Q9x909 streptomyce
17	80	37.6	194	1	H1_SALTR	P02254 salmo trutt
18	80	37.6	206	1	H1_ONCMY	P06350 oncorhynchu
19	79.5	37.3	217	1	H1G_STRPU	P07796 strongyloce
20	79	37.1	244	1	H1O_CHITH	Q07134 chironomus
21	78.5	36.9	236	1	H1_NEUCR	Q8j0u2 neurospora
22	78.5	36.9	237	1	H1E_CHITE	P40278 chironomus
23	78	36.6	233	1	H11_GLYSA	P40266 glyptotendi
24	78	36.6	287	1	H1_LYCES	P37218 lycopersico
25	77.5	36.4	299	1	RL22_DROME	P50887 drosophila
26	77	36.2	219	1	H110_CHICK	P08286 gallus gall
27	77	36.2	225	1	H15_HUMAN	P16401 homo sapien
28	76.5	35.9	139	1	ASR_KLEPN	Q93mh5 klebsiella
29	76.5	35.9	232	1	H12_GLYBA	P40264 glyptotendi
30	76.5	35.9	233	1	H11_GLYBA	P40263 glyptotendi
31	76.5	35.9	238	1	H1_WHEAT	P27806 triticum ae
32	76	35.7	240	1	H12_VOLCA	Q08865 volvox cart
33	75.5	35.4	184	1	RS16_BACTN	Q9rq15 bacteroides
34	75.5	35.4	212	1	RL14_HUMAN	P50914 homo sapien
35	75	35.2	211	1	H12_MOUSE	P15864 mus musculu
36	74.5	35.0	101	1	ASR_ENTCL	Q93mh6 enterobacte
37	74.5	35.0	165	1	RS16_CAUCR	P58122 caulobacter
38	74.5	35.0	1002	1	IF2P_YEAST	P39730 saccharomyc
39	74	34.7	153	1	RS16_BIFLO	Q8g7g1 bifidobacte
40	74	34.7	352	1	ALGP_PSEAE	P15276 pseudomonas
41	74	34.7	668	1	FAU_DROME	Q9vgx3 drosophila
42	73	34.3	153	1	H2B1_CHLRE	P50565 chlamydomon
43	73	34.3	190	1	H12_CAEEL	P15796 caenorhabdi
44	73	34.3	207	1	H11_CAEEL	P10771 caenorhabdi
45	73	34.3	220	1	H13_MOUSE	P43277 mus musculu

ALIGNMENTS

RESULT 1
H1_LYTPI
ID H1_LYTPI STANDARD; PRT; 210 AA.
AC P06144;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Late histone H1.
OS Lytechinus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae;
OC Lytechinus.
OX NCBI_TaxID=7653;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RT Lytechinus pictus and Strongelocentrotus purpuratus.";
RL Nucleic Acids Res. 14:8121-8133(1986).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC ---

CC ---
DR EMBL; X04488; CAA28177.1; -.
DR PIR; A25550; A25550.
DR HSSP; P02259; 1HST.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker_histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;

Query Match 48.6%; Score 103.5; DB 1; Length 210;
Best Local Similarity 65.1%; Pred. No. 0.0011;
Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

```
Qy          1 AKKYAKKAKAEK-AKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              ||| |||  |:| |||| |   ||||||  | ||:| |||:||
Db        166 AKKAAKKPAAKKPAKKAAKKPAAKKAAKPAKKAAKKPAAKKAA 208
```

```
RESULT 2
TOLA_ECOLI
ID   TOLA_ECOLI     STANDARD;      PRT;   421 AA.
AC   P19934;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   TolA protein.
GN   TOLA OR CIM OR EXCC OR LKY OR B0739.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JM105;
RX   MEDLINE=90078104; PubMed=2687247;
RA   Levengood S.K., Webster R.E.;
RT   "Nucleotide sequences of the tolA and tolB genes and localization of
RT   their products, components of a multistep translocation system in
RT   Escherichia coli.";
RL   J. Bacteriol. 171:6600-6609(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   DOMAINS.
RX   MEDLINE=91296736; PubMed=2068069;
RA   Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT   "TolA: a membrane protein involved in colicin uptake contains an
RT   extended helical region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN   [5]
RP   INTERACTION WITH PORINS.
RX   MEDLINE=97133271; PubMed=8978668;
RA   Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA   Lloubes R.;
RT   "TolA central domain interacts with Escherichia coli porins.";
RL   EMBO J. 15:6408-6415(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX   MEDLINE=99332679; PubMed=10404600;
RA   Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT   "Filamentous phage infection: crystal structure of g3p in complex
RT   with its coreceptor, the C-terminal domain of TolA.";
RL   Structure 7:711-722(1999).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC       COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC       COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC       BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC       OF BACTERIOPHAGE DNA.
CC   -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC       AND LAMB.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28232; AAA24683.1; -.
DR   EMBL; AE000177; AAC73833.1; -.
DR   EMBL; D90713; BAA35405.1; -.
DR   PIR; JV0057; JV0057.
DR   PDB; 1TOL; 20-MAY-99.
DR   EcoGene; EG11007; tolA.
KW   Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW   Repeat; Inner membrane; 3D-structure; Complete proteome.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      14    34       POTENTIAL.
FT   DOMAIN        35    421      PERIPLASMIC (POTENTIAL).
FT   DOMAIN        48    310      DOMAIN II (ALPHA-HELICAL).
FT   DOMAIN        311   421      DOMAIN III (FUNCTIONAL).
FT   DOMAIN        224   278      10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT                                A(2,4).
FT   HELIX         335   349
FT   TURN          350   351
FT   TURN          353   354
FT   HELIX         355   358
FT   TURN          359   360
FT   STRAND        363   369
FT   TURN          371   372
FT   STRAND        375   383
FT   HELIX         385   397
FT   HELIX         406   412
FT   TURN          413   414
FT   STRAND        416   421
SQ   SEQUENCE   421 AA;  43156 MW;  8B2F52B4B97C655E CRC64;

  Query Match             46.0%;  Score 98;  DB 1;  Length 421;
  Best Local Similarity   65.3%;  Pred. No. 0.0073;
  Matches   32;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || :|| ||||   ||  | |||  ||||   ||||||||| :|| ||
Db        247 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 294


RESULT 3
MST2_DROHY
ID   MST2_DROHY     STANDARD;      PRT;  1391 AA.
AC   Q08696;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Axoneme-associated protein mst101(2).
GN   MST101(2).
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=95045538; PubMed=7957199;
RA   Neesen J., Padmanabhan S., Buenemann H.;
RT   "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT   motif representing the major component of the sperm-tail-specific
RT   axoneme-associated protein family Dhmst101 form extended
RT   alpha-helical rods within the extremely elongated spermatozoa of
RT   Drosophila hydei.";
RL   Eur. J. Biochem. 225:1089-1095(1994).
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
```

```
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC       SPERMATIDS.
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC   -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC       STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC       REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73481; CAA51876.1; -.
DR   PIR; S51364; S51364.
DR   FlyBase; FBgn0020733; Dhyd\mst101(2).
KW   Sperm; Repeat; Multigene family; Polymorphism.
FT   DOMAIN      332   1268       59 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                [KR]-K-X-C-X-X-X-A-K-X-X-K-X-X-X-E.
SQ   SEQUENCE   1391 AA;  159000 MW;  1B2A368F30F48878 CRC64;

  Query Match             42.3%;  Score 90;  DB 1;  Length 1391;
  Best Local Similarity   57.8%;  Pred. No. 0.13;
  Matches   26;  Conservative    4;  Mismatches   11;  Indels    4;  Gaps    2;

Qy          2 KKYAKKAKAEK-AKKAYKAAEAKK---AAKYEKAAAEKAAAKEAA 42
              |: |:| | || |||  :||| ||   ||| || ||||   :|||
Db        663 KETAEKKKCEKAAKKRKEAAEKKKCAEAAKKEKEAAEKKKCEEAA 707


RESULT 4
DBH_MYCSM
ID   DBH_MYCSM      STANDARD;      PRT;   208 AA.
AC   Q9ZHC5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN   HUP OR HLP.
OS   Mycobacterium smegmatis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 700084 / mc(2)155;
RX   MEDLINE=99110209; PubMed=9894918;
RA   Lee B.H., Murugasu-Oei B., Dick T.;
RT   "Upregulation of a histone-like protein in dormant Mycobacterium
RT   smegmatis.";
RL   Mol. Gen. Genet. 260:475-479(1998).
CC   -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC       PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC       DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC       ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068138; AAD13809.1; -.
DR   HSSP; P02346; 1HUU.
DR   InterPro; IPR000119; Bac_DNAbind.
DR   Pfam; PF00216; Bac_DNA_binding; 1.
DR   ProDom; PD000945; Bac_DNAbind; 1.
DR   SMART; SM00411; BHL; 1.
DR   PROSITE; PS00045; HISTONE_LIKE; 1.
KW   DNA-binding; DNA condensation; Repeat.
FT   DOMAIN        1     90       BACTERIAL HISTONE-LIKE DOMAIN.
FT   DOMAIN      101    205       DEGENERATE REPEATS REGION.
SQ   SEQUENCE   208 AA;  21230 MW;  CA5F577F61F7EF09 CRC64;

  Query Match             42.0%;  Score 89.5;  DB 1;  Length 208;
  Best Local Similarity   53.7%;  Pred. No. 0.025;
  Matches   29;  Conservative    3;  Mismatches   13;  Indels    9;  Gaps    2;

Qy          1 AKKYAKKAKAEKA-------KKAYKAAEAKKAAKY--EKAAAEKAAAKEAAYEA 45
              ||| |||| |:||       | | | | |||||     | || || ||:|| :|
Db        111 AKKAAKKAPAKKAAAKKTATKAAAKKAPAKKAATKAPAKKAATKAPAKKAATKA 164


RESULT 5
H1B_STRPU
ID   H1B_STRPU      STANDARD;      PRT;   211 AA.
AC   P15869;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1-beta, late embryonic.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88246461; PubMed=2837660;
RA   Lai Z.-C., Childs G.;
RT   "Characterization of the structure and transcriptional patterns of
RT   the gene encoding the late histone subtype H1-beta of the sea urchin
RT   Strongylocentrotus purpuratus.";
RL   Mol. Cell. Biol. 8:1842-1844(1988).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20314; AAA30052.1; -.
DR   PIR; A28100; A28100.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; LinkerHist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   211 AA;  22169 MW;  9F214581334BBE7A CRC64;

  Query Match             42.0%;  Score 89.5;  DB 1;  Length 211;
  Best Local Similarity   61.9%;  Pred. No. 0.026;
  Matches   26;  Conservative    3;  Mismatches   10;  Indels    3;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              :||  || |   |||| | | ||||||  | ||:| |||:||
Db        139 SKKTTKKVKKPAAKKAKKPA-AKKAAK--KPAAKKPAAKKAA 177


RESULT 6
```

```
DBH_MYCTU
ID   DBH_MYCTU      STANDARD;      PRT;   214 AA.
AC   P95109;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa
DE   laminin-2-binding protein).
GN   HUP OR HLP OR LBP21 OR RV2986C OR MT3064 OR MTCY349.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 71-86, AND DNA-BINDING.
RC   STRAIN=H37Rv;
RA   Prasad H.K., Annapurna P.S., Dey A.B., Tyagi J.S., Jain N.K.,
RA   Savita P.;
RL   Submitted (DEC-1997) to the SWISS-PROT data bank.
CC   -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC       PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC       DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC       ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z83018; CAB05427.1; -.
DR   EMBL; AE007127; AAK47393.1; -.
DR   PIR; G70673; G70673.
DR   HSSP; P02346; 1HUU.
DR   TIGR; MT3064; -.
DR   TubercuList; Rv2986c; -.
DR   InterPro; IPR000119; Bac_DNAbind.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00216; Bac_DNA_binding; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000945; Bac_DNAbind; 1.
DR   SMART; SM00411; BHL; 1.
DR   PROSITE; PS00045; HISTONE_LIKE; 1.
KW   DNA-binding; DNA condensation; Repeat; Complete proteome.
FT   DOMAIN        1     90       BACTERIAL HISTONE-LIKE DOMAIN.
FT   DOMAIN      101    214       DEGENERATE REPEATS REGION.
SQ   SEQUENCE   214 AA;  22187 MW;  CB09AF20FB353544 CRC64;

  Query Match             42.0%;  Score 89.5;  DB 1;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.026;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  | |   : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 7
TMPB_TREPH
ID   TMPB_TREPH     STANDARD;      PRT;   384 AA.
AC   P29720;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Treponemal membrane protein B precursor (Antigen tmpB).
GN   TMPB.
OS   Treponema phagedenis.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kazan 5;
RX   MEDLINE=91372983; PubMed=1894368;
RA   Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,
RA   Sliviensky L., Schouls L.M., van Embden J.D., Charon N.W.;
RT   "Treponema phagedenis encodes and expresses homologs of the Treponema
RT   pallidum TmpA and TmpB proteins.";
RL   Infect. Immun. 59:3685-3693(1991).
CC   -!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR
CC       LARGE MOLECULES.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.
CC   -!- SIMILARITY: TO TMPB OF T.PALLIDUM.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M58563; AAA27480.1; -.
DR   PIR; B43592; B43592.
KW   Antigen; Outer membrane; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    384       TREPONEMAL MEMBRANE PROTEIN B.
FT   DOMAIN      151    235       17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR]-
FT                                [ED].
FT   REPEAT      151    155       1-1.
FT   REPEAT      156    160       1-2.
FT   REPEAT      161    165       1-3.
FT   REPEAT      166    170       1-4.
FT   REPEAT      171    175       1-5.
FT   REPEAT      176    180       1-6.
FT   REPEAT      181    185       1-7.
FT   REPEAT      186    190       1-8.
FT   REPEAT      191    195       1-9.
FT   REPEAT      196    200       1-10.
FT   REPEAT      201    205       1-11.
FT   REPEAT      206    210       1-12.
FT   REPEAT      211    215       1-13.
FT   REPEAT      216    220       1-14.
FT   REPEAT      221    225       1-15.
FT   REPEAT      226    230       1-16.
FT   REPEAT      231    235       1-17.
FT   DOMAIN      236    288       6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-
FT                                A-A-E.
```

```
FT   REPEAT       236    243       2-1.
FT   REPEAT       245    252       2-2.
FT   REPEAT       254    261       2-3.
FT   REPEAT       263    270       2-4.
FT   REPEAT       272    279       2-5.
FT   REPEAT       281    288       2-6.
SQ   SEQUENCE   384 AA;  42677 MW;  6E94CBC74294DE8C CRC64;

  Query Match             40.6%;  Score 86.5;  DB 1;  Length 384;
  Best Local Similarity   54.5%;  Pred. No. 0.086;
  Matches   24;  Conservative    6;  Mismatches   13;  Indels    1;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYE 44
              ||: | | || | | | : |  :|||| :||| |||| ::|| |
Db        178 AKEKAAKEKAAKDKAAKEKAAKEKAAK-DKAAKEKAAKEKAARE 220


RESULT 8
RS16_COREF
ID   RS16_COREF     STANDARD;      PRT;   168 AA.
AC   Q8FP30;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   30S ribosomal protein S16.
GN   RPSP OR CE1960.
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA   Usuda Y., Sugimoto S.;
RT   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005220; BAC18770.1; -.
DR   HAMAP; MF_00385; -; 1.
DR   InterPro; IPR000307; Ribosomal_S16.
DR   Pfam; PF00886; Ribosomal_S16; 1.
DR   ProDom; PD003791; Ribosomal_S16; 1.
DR   TIGRFAMs; TIGR00002; S16; 1.
DR   PROSITE; PS00732; RIBOSOMAL_S16; FALSE_NEG.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   168 AA;  18228 MW;  443CE68AA2321C01 CRC64;

  Query Match             40.4%;  Score 86;  DB 1;  Length 168;
  Best Local Similarity   65.7%;  Pred. No. 0.046;
  Matches   23;  Conservative    3;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          7 KAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              :|  || ||| :  |||:||  |||||||||| ||
Db        120 EAITEKKKKAREEKEAKEAA--EKAAAEKAAAAEA 152


RESULT 9
HBHA_MYCTU
ID   HBHA_MYCTU     STANDARD;      PRT;   198 AA.
AC   Q11142; O85733;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Heparin-binding hemagglutinin (Adhesin).
GN   HBHA OR RV0475 OR MT0493 OR MTCY20G9.01.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC   SPECIES=M.tuberculosis, and M.bovis;
RC   STRAIN=H37Rv, ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX   MEDLINE=98445421; PubMed=9770536;
RA   Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Locht C.;
RT   "Molecular characterization of the mycobacterial heparin-binding
RT   hemagglutinin, a mycobacterial adhesin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12625-12630(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-16, AND CHARACTERIZATION.
RC   SPECIES=M.tuberculosis, and M.bovis;
RC   STRAIN=ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX   MEDLINE=97188915; PubMed=9064359;
RA   Menozzi F.D., Rouse J.H., Alavi M., Laude-Sharp M., Muller J.,
RA   Bischoff R., Brennan M.J., Locht C.;
RT   "Identification of a heparin-binding hemagglutinin present in
RT   Mycobacteria.";
RL   J. Exp. Med. 184:993-1001(1996).
RN   [5]
RP   FUNCTION.
RC   SPECIES=M.tuberculosis, and M.bovis;
RC   STRAIN=103, and BCG;
RX   MEDLINE=21342355; PubMed=11449276;
RA   Pethe K., Alonso S., Biet F., Delogu G., Brennan M.J., Locht C.,
RA   Menozzi F.D.;
RT   "The heparin-binding haemagglutinin of M. tuberculosis is required for
RT   extrapulmonary dissemination.";
RL   Nature 412:190-194(2001).
CC   -!- FUNCTION: REQUIRED FOR EXTRAPULMONARY DISSEMINATION. MEDIATES
CC       ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED
CC       GLYCOCONJUGATES PRESENT AT THE SURFACE OF THESE CELLS; BINDS
CC       HEPARIN, DEXTRAN SULFATE, FUCOIDAN AND CHONDROITIN SULFATE.
CC       PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.
CC       INDUCES MYCOBACTERIAL AGGREGATION.
CC   -!- SUBCELLULAR LOCATION: SURFACE ASSOCIATED.
CC   -!- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL
CC       DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END
CC       DIMINISH THE AFFINITY FOR HEPARIN.
```

```
CC   -!- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
CC       PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
CC       SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
CC       TERMINAL DOMAIN OF HBHA.
CC   -!- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
CC       TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
CC       RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CC       CONTAIN ANY.
CC   -!- SIMILARITY: STRONG, TO M.LEPRAE HBHA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF074390; AAC26052.1; -.
DR   EMBL; Z77162; CAB00936.1; -.
DR   EMBL; AE006951; AAK44716.1; ALT_INIT.
DR   PIR; F70742; F70742.
DR   TIGR; MT0493; -.
DR   TubercuList; Rv0475; -.
KW   Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
KW   Virulence; Complete proteome.
FT   INIT_MET      0      0
FT   DOMAIN      151    193       ALA/LYS-RICH.
FT   CONFLICT    120    120       R -> P (IN REF. 1).
SQ   SEQUENCE   198 AA;  21403 MW;  513760F6F1EB6042 CRC64;

  Query Match             40.4%;  Score 86;  DB 1;  Length 198;
  Best Local Similarity   65.6%;  Pred. No. 0.053;
  Matches   21;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         11 EKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              :||  | ||| |||||  :||||:|| ||:||
Db        160 KKAAPAKKAAPAKKAAPAKKAAAKKAPAKKAA 191


RESULT 10
MST1_DROHY
ID   MST1_DROHY     STANDARD;      PRT;   344 AA.
AC   Q08695;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Axoneme-associated protein mst101(1).
GN   MST101(1).
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=94200512; PubMed=8150205;
RA   Neesen J., Buenemann H., Heinlein U.A.;
RT   "The Drosophila hydei gene Dhmst101(1) encodes a testis-specific,
RT   repetitive, axoneme-associated protein with differential abundance in
RT   Y chromosomal deletion mutant flies.";
RL   Dev. Biol. 162:414-425(1994).
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC       ASSOCIATED WITH AXONEMAL STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC       SPERMATID BUNDLES.
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73480; CAA51875.1; -.
DR   PIR; S34153; S34153.
DR   FlyBase; FBgn0011816; Dhyd\mst101(1).
KW   Sperm; Repeat; Multigene family.
FT   DOMAIN       58    337       19 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.
SQ   SEQUENCE   344 AA;  37793 MW;  24C65D2510387E2A CRC64;

  Query Match             40.4%;  Score 86;  DB 1;  Length 344;
  Best Local Similarity   57.8%;  Pred. No. 0.087;
  Matches   26;  Conservative    3;  Mismatches   12;  Indels    4;  Gaps    2;

Qy          2 KKYAKKAK-AEKAKKAYKAAEAKK---AAKYEKAAAEKAAAKEAA 42
              |: |:| | || |||  :||| ||   ||| || ||||    |||
Db         69 KEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAA 113


RESULT 11
TOLA_PSEAE
ID   TOLA_PSEAE     STANDARD;      PRT;   347 AA.
AC   P50600;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TolA protein.
GN   TOLA OR PA0971.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO;
RX   MEDLINE=97113525; PubMed=8955385;
RA   Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT   "Identification and characterization of the tolQRA genes of
RT   Pseudomonas aeruginosa.";
RL   J. Bacteriol. 178:7059-7068(1996).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RA   Duan K., Sokol P.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC       (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; U39558; AAC44660.2; -.
DR   EMBL; AE004530; AAG04360.1; -.
DR   PIR; E83525; E83525.
DR   InterPro; IPR006260; TonB_C.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
KW   Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW   Complete proteome.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     37       POTENTIAL.
FT   DOMAIN       38    347       PERIPLASMIC (POTENTIAL).
FT   DOMAIN      209    216       POLY-ALA.
SQ   SEQUENCE   347 AA;  37935 MW;  EEDD4B04AA095945 CRC64;

  Query Match             40.4%;  Score 86;  DB 1;  Length 347;
  Best Local Similarity   51.1%;  Pred. No. 0.088;
  Matches   24;  Conservative    6;  Mismatches   15;  Indels    2;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKA--YKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |:| |:  ||::|||   |||| || |   |  ||  | |:|| :|
Db        125 AQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKRAEDEAKKKAAEDA 171


RESULT 12
DBH_MYCBO
ID   DBH_MYCBO      STANDARD;      PRT;   205 AA.
AC   Q9XB18; Q9S5J5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN   HUP OR HLP OR MDP1.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AN5;
RA   Prabhakar S., Tyagi J.S., Prasad H.K.;
RT   "HLPMt-A target for differentiation of M.tuberculosis and M.bovis.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BCG / Tokyo;
RA   Matsumoto S., Yukitake H., Matsuo T., Mineda T., Yamada T.;
RT   "Identification of a novel protein generating bacterial slow growth
RT   from Mycobacterium.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC       PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC       DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC       ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y18421; CAB46493.1; -.
DR   EMBL; AB013441; BAA78330.1; -.
DR   HSSP; P02346; 1HUU.
DR   InterPro; IPR000119; Bac_DNAbind.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00216; Bac_DNA_binding; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000945; Bac_DNAbind; 1.
DR   SMART; SM00411; BHL; 1.
DR   PROSITE; PS00045; HISTONE_LIKE; 1.
KW   DNA-binding; DNA condensation; Repeat.
FT   DOMAIN        1     90       BACTERIAL HISTONE-LIKE DOMAIN.
FT   DOMAIN      101    205       DEGENERATE REPEATS REGION.
FT   CONFLICT    199    199       A -> T (IN REF. 2).
SQ   SEQUENCE   205 AA;  21262 MW;  19FCE67885DFE6A8 CRC64;

  Query Match             40.1%;  Score 85.5;  DB 1;  Length 205;
  Best Local Similarity   58.5%;  Pred. No. 0.061;
  Matches   24;  Conservative    2;  Mismatches   12;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              ||| |||| |:|| || | |  |  |   | || || ||:|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---KKAATKAPAKKA 148


RESULT 13
RS16_CORGL
ID   RS16_CORGL     STANDARD;      PRT;   165 AA.
AC   Q8NNX3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S16.
GN   RPSP OR CGL2054.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP005280; BAB99447.1; -.
DR   HAMAP; MF_00385; -; 1.
DR   InterPro; IPR000307; Ribosomal_S16.
DR   Pfam; PF00886; Ribosomal_S16; 1.
DR   ProDom; PD003791; Ribosomal_S16; 1.
DR   TIGRFAMs; TIGR00002; S16; 1.
DR   PROSITE; PS00732; RIBOSOMAL_S16; FALSE_NEG.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   165 AA;  17837 MW;  61DD81961BC30846 CRC64;

  Query Match             38.5%;  Score 82;  DB 1;  Length 165;
  Best Local Similarity   62.9%;  Pred. No. 0.11;
  Matches   22;  Conservative    4;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          7 KAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              :|  || ||| :  |||:||  |||||||||| |:
Db        120 EAITEKKKKAREDKEAKEAA--EKAAAEKAAAAES 152


RESULT 14
H1_PARAN
ID   H1_PARAN       STANDARD;      PRT;   248 AA.
AC   P02256;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1, gonadal.
OS   Parechinus angulosus (Angulate sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
```

```
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
OC   Parechinus.
OX   NCBI_TaxID=7658;
RN   [1]
RP   SEQUENCE OF 1-84.
RX   MEDLINE=80156831; PubMed=6767609;
RA   Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT   protein and the sequence of amino acids in the four N-terminal
RT   cyanogen bromide peptides.";
RL   Eur. J. Biochem. 104:559-566(1980).
RN   [2]
RP   SEQUENCE OF 80-248.
RX   MEDLINE=80156832; PubMed=7363905;
RA   Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT   the entire primary structure.";
RL   Eur. J. Biochem. 104:567-578(1980).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: SPERM.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR   PIR; A91090; HSUR1P.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Sperm.
FT   VARIANT     144    144       K -> R.
SQ   SEQUENCE   248 AA;  26387 MW;  1B25B3F136541947 CRC64;

  Query Match             38.5%;  Score 82;  DB 1;  Length 248;
  Best Local Similarity   57.8%;  Pred. No. 0.16;
  Matches   26;  Conservative    3;  Mismatches   14;  Indels    2;  Gaps    2;

Qy          2 KKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAA-AKEAAYEA 45
              ||  | : | ||||| ||| |||| | : ||  ||| ||: |  |
Db        120 KKAKKTSAAAKAKKA-KAAAAKKARKAKAAAKRKAALAKKKAAAA 163


RESULT 15
H1E_CHIPA
ID   H1E_CHIPA      STANDARD;      PRT;   235 AA.
AC   P40262;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1E.
OS   Chironomus pallidivittatus (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Chironomidae; Chironominae; Chironomus.
OX   NCBI_TaxID=7151;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulze E., Wisniewski J.R., Nagel S., Gavenis K., Grossbach U.;
RL   Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29106; AAA21714.1; -.
DR   HSSP; P08287; 1GHC.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   235 AA;  24446 MW;  340BC5B9A85002AC CRC64;

  Query Match             38.0%;  Score 81;  DB 1;  Length 235;
  Best Local Similarity   41.3%;  Pred. No. 0.19;
  Matches   26;  Conservative    2;  Mismatches   17;  Indels   18;  Gaps    1;

Qy          1 AKKYAKKAKAEKA----------------KKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              |||  |  ||  |                | | |  ||||| |  | ||:| ||| ||
Db        126 AKKVDKPKKAPAAPKPKSTNKRVTGEKKVVKKPAAKKPEAKKATKAAKPAAKKVAAKPAA 185

Qy         43 YEA 45
                :|
Db        186 KKA 188


Search completed: January 21, 2004, 09:00:58
Job time : 8.13636 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:53:31 ; Search time 18.1169 Seconds
(without alignments)
640.969 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	98	46.0	394	16	Q8X965	Q8x965 escherichia
2	98	46.0	421	16	Q8FJT1	Q8fjt1 escherichia
3	94.5	44.4	239	16	Q8Y5W4	Q8y5w4 listeria mo
4	91	42.7	244	16	Q9AJX2	Q9ajx2 streptomyce
5	90	42.3	243	16	Q92A67	Q92a67 listeria in
6	88.5	41.5	275	5	O01395	O01395 drosophila
7	88.5	41.5	372	2	Q9WWX1	Q9wwx1 pseudomonas
8	88	41.3	899	16	Q9A5J6	Q9a5j6 caulobacter
9	86.5	40.6	200	16	Q8XVN7	Q8xvn7 ralstonia s
10	86	40.4	155	16	Q8PI40	Q8pi40 xanthomonas
11	86	40.4	168	16	Q8FP30	Q8fp30 corynebacte
12	86	40.4	212	3	O93946	O93946 candida alb
13	85	39.9	98	5	Q8WQ44	Q8wq44 leishmania
14	85	39.9	111	5	Q8T9R3	Q8t9r3 leishmania
15	85	39.9	182	2	Q45370	Q45370 bordetella
16	84	39.4	60	5	Q9U3W3	Q9u3w3 caenorhabdi
17	84	39.4	144	13	Q8AWI3	Q8awi3 oncorhynchu
18	83.5	39.2	483	12	Q8QND1	Q8qnd1 ectocarpus
19	83	39.0	293	10	Q9AT18	Q9at18 lens culina
20	83	39.0	389	16	Q9CM70	Q9cm70 pasteurella
21	82.5	38.7	243	5	Q23784	Q23784 chironomus
22	82	38.5	190	5	Q8MYC3	Q8myc3 mytilus cal
23	82	38.5	262	16	Q9S2M2	Q9s2m2 streptomyce
24	82	38.5	568	3	O94567	O94567 schizosacch
25	81.5	38.3	156	16	Q8P6U0	Q8p6u0 xanthomonas
26	81.5	38.3	311	12	Q84528	Q84528 paramecium
27	81.5	38.3	482	2	Q93LK4	Q93lk4 enterococcu
28	81	38.0	81	5	Q9NFP6	Q9nfp6 trypanosoma
29	81	38.0	112	5	Q9XYY5	Q9xyy5 leishmania
30	81	38.0	241	5	Q23790	Q23790 chironomus
31	81	38.0	290	10	Q9AT24	Q9at24 pisum sativ
32	81	38.0	295	10	Q8LKH9	Q8lkh9 pisum fulvu
33	81	38.0	295	10	Q8LKI0	Q8lki0 pisum sativ
34	81	38.0	295	10	Q9AT22	Q9at22 lathyrus sa
35	81	38.0	295	10	Q9ZR20	Q9zr20 pisum sativ
36	81	38.0	296	10	Q9AT25	Q9at25 pisum sativ
37	81	38.0	297	10	Q9SXQ8	Q9sxq8 pisum sativ
38	81	38.0	298	10	Q8LKI1	Q8lki1 lathyrus ap
39	81	38.0	301	10	Q9AT23	Q9at23 pisum sativ
40	81	38.0	306	10	Q9AT21	Q9at21 lathyrus sa
41	80.5	37.8	277	16	Q9XAQ3	Q9xaq3 streptomyce
42	80.5	37.8	445	3	Q9P3Q8	Q9p3q8 neurospora
43	80	37.6	66	5	Q95QZ0	Q95qz0 caenorhabdi
44	80	37.6	71	5	Q9NFJ8	Q9nfj8 trypanosoma
45	80	37.6	81	5	Q9N6L9	Q9n6l9 trypanosoma

ALIGNMENTS

RESULT 1
Q8X965
ID Q8X965 PRELIMINARY; PRT; 394 AA.
AC Q8X965;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Membrane spanning protein, required for outer membrane integrity
DE (Membrane spanning protein TolA).
GN TOLA OR Z0907 OR ECS0774.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005252; AAG55075.1; -.
DR EMBL; AP002553; BAB34197.1; -.

```
DR   InterPro; IPR000104; Antifreeze_1.
DR   PRINTS; PR00308; ANTIFREEZEI.
KW   Complete proteome.
SQ   SEQUENCE   394 AA;  40517 MW;  5B58D8E8230BDE28 CRC64;

  Query Match             46.0%;  Score 98;  DB 16;  Length 394;
  Best Local Similarity   65.3%;  Pred. No. 0.033;
  Matches   32;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || |:|| ||||   ||  | |||  ||||   ||||||||| :|| ||
Db        220 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 267


RESULT 2
Q8FJT1
ID   Q8FJT1      PRELIMINARY;      PRT;   421 AA.
AC   Q8FJT1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TolA protein.
GN   TOLA OR C0818.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016757; AAN79291.1; -.
KW   Complete proteome.
SQ   SEQUENCE   421 AA;  43184 MW;  DB296626F056D385 CRC64;

  Query Match             46.0%;  Score 98;  DB 16;  Length 421;
  Best Local Similarity   65.3%;  Pred. No. 0.035;
  Matches   32;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAKAEKA---KK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              || |:|| ||||   ||  | |||  ||||   ||||||||| :|| ||
Db        247 KKAAEKAAAEKAAADKKAAAEKAAADKKAAA-AKAAAEKAAAAKAAAEA 294


RESULT 3
Q8Y5W4
ID   Q8Y5W4      PRELIMINARY;      PRT;   239 AA.
AC   Q8Y5W4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein lmo1941.
GN   LMO1941.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591981; CAD00019.1; -.
DR   ListiList; LMO01941; -.
DR   InterPro; IPR002482; LysM.
DR   Pfam; PF01476; LysM; 1.
DR   SMART; SM00257; LysM; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   239 AA;  25836 MW;  72E59D576E0D7832 CRC64;

  Query Match             44.4%;  Score 94.5;  DB 16;  Length 239;
  Best Local Similarity   56.2%;  Pred. No. 0.045;
  Matches   27;  Conservative    7;  Mismatches    9;  Indels    5;  Gaps    2;

Qy          2 KKYAKKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAKEAAYE 44
              |  |:||:|:| |:   | ||| |||   |  |||||:||||::|| |
Db        124 KAAAEKAEADKKKQEEDAVKAANAKKEQEAAEEKAAADKAAAEKAAAE 171


RESULT 4
Q9AJX2
ID   Q9AJX2      PRELIMINARY;      PRT;   244 AA.
AC   Q9AJX2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative secreted protein.
GN   SCO1805 OR SCI33.04.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL939110; CAC28545.1; -.
KW   Complete proteome.
SQ   SEQUENCE   244 AA;  25524 MW;  61999D62CA23A7B0 CRC64;

  Query Match             42.7%;  Score 91;  DB 16;  Length 244;
  Best Local Similarity   59.5%;  Pred. No. 0.1;
  Matches   25;  Conservative    4;  Mismatches   11;  Indels    2;  Gaps    1;

Qy          5 AKKAKAEKAKK--AYKAAEAKKAAKYEKAAAEKAAAKEAAYE 44
              ::|  | :|||  | | | ||||   :||||||||| || |
Db         78 SQKVAAIEAKKEAAAKKATAKKATAEKKAAAEKAAAKRAAKE 119


RESULT 5
Q92A67
ID   Q92A67      PRELIMINARY;      PRT;   243 AA.
AC   Q92A67;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
DE   Hypothetical protein lin2055.
GN   LIN2055.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / Serovar 6a;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596170; CAC97285.1; -.
DR   ListiList; LIN02055; -.
DR   InterPro; IPR002482; LysM.
DR   Pfam; PF01476; LysM; 1.
DR   SMART; SM00257; LysM; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   243 AA;  25963 MW;  6B2493D143B159D1 CRC64;

  Query Match             42.3%;  Score 90;  DB 16;  Length 243;
  Best Local Similarity   45.5%;  Pred. No. 0.13;
  Matches   30;  Conservative    3;  Mismatches   11;  Indels   22;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKA----------------------AKYEKAAAEKAAA 38
              ||| | :|| |||| :  |||| | |                      |  |||||||||
Db        106 AKKAAEEKAAAEKAAEEKKAAEEKAAADKKSQEDEAAKAAAAKKEQEAAEEKAAAEKAAA 165

Qy         39 KEAAYE 44
                :|| |
Db        166 DKAAKE 171


RESULT 6
O01395
ID   O01395      PRELIMINARY;      PRT;   275 AA.
AC   O01395;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Axoneme-associated protein MST101(3).
GN   MST101(3) OR DHMST101.
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Neesen J., Heinlein U.A.O., Buenemann H.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR   EMBL; U85627; AAB51369.1; -.
DR   FlyBase; FBgn0020732; Dhyd\mst101(3).
KW   Sperm; Repeat; Multigene family.
FT   DOMAIN       64    255       13 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                X-[KQ]]-K-C-[AE]-E-X-A-[X]-K-X-X-X-X-
FT                                [AE]-X.
SQ   SEQUENCE   275 AA;  30436 MW;  76BAA7B2A2DF732C CRC64;

  Query Match             41.5%;  Score 88.5;  DB 5;  Length 275;
  Best Local Similarity   61.0%;  Pred. No. 0.2;
  Matches   25;  Conservative    1;  Mismatches   12;  Indels    3;  Gaps    1;

Qy          5 AKKAKAEKAKKAYKAAEAKK---AAKYEKAAAEKAAAKEAA 42
              |||  || |||  :||| ||   ||| || ||||    |||
Db         63 AKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAA 103


RESULT 7
Q9WWX1
ID   Q9WWX1      PRELIMINARY;      PRT;   372 AA.
AC   Q9WWX1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TolA protein.
GN   TOLA.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RX   MEDLINE=96198174; PubMed=8626299;
RA   Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;
RT   "The Pseudomonas putida peptidoglycan-associated outer membrane
RT   lipoprotein (PAL) is involved in maintenance of the integrity of the
RT   cell envelope.";
RL   J. Bacteriol. 178:1699-1706(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RA   Ramos-Gonzalez I.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RA   Rodriguez-Herva J.J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RX   MEDLINE=96422022; PubMed=8824639;
RA   Rodriguez-Herva J.J., Ramos J.;
RT   "Characterization of an OprL null mutant of Pseudomonas putida.";
RL   J. Bacteriol. 178:5836-5840(1996).
DR   EMBL; X74218; CAB50780.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR006260; TonB_C.
DR   PRINTS; PR00624; HISTONEH5.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
SQ   SEQUENCE   372 AA;  40133 MW;  87F49785ECC3C0BC CRC64;

  Query Match             41.5%;  Score 88.5;  DB 2;  Length 372;
  Best Local Similarity   39.7%;  Pred. No. 0.27;
  Matches   27;  Conservative    8;  Mismatches   10;  Indels   23;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAA----------------------AEKAA 37
              |:  || |:| || :| |||||||| : :|||                      |:|||
Db        121 AEDAAKAAEAAKAAEAKKAAEAKKADEAKKAAEKQQADIAKKKAEDEAKKKAEEEAKKAA 180

Qy         38 AKEAAYEA 45
                |:||  :|
Db        181 AEEAKKKA 188


RESULT 8
Q9A5J6
ID   Q9A5J6      PRELIMINARY;      PRT;   899 AA.
AC   Q9A5J6;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DNA topoisomerase (EC 5.99.1.2).
GN   CC2451.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC       CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC       DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC   -!- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC       BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC       WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC       AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC       FAMILY.
DR   EMBL; AE005914; AAK24422.1; -.
DR   HSSP; P06612; 1ECL.
DR   TIGR; CC2451; -.
DR   InterPro; IPR003601; DNAtopI_ATP_bind.
DR   InterPro; IPR003602; DNAtopI_DNA_bind.
DR   InterPro; IPR005733; DNA_topI_bact.
DR   InterPro; IPR000380; DNA_tpisomrase.
DR   InterPro; IPR006171; Toprim_dom.
DR   InterPro; IPR006154; Toprim_sub.
DR   Pfam; PF01131; Topoisom_bac; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   Pfam; PF01396; zf-C4_Topoisom; 1.
DR   PRINTS; PR00417; PRTPISMRASEI.
DR   SMART; SM00437; TOP1Ac; 1.
DR   SMART; SM00436; TOP1Bc; 1.
DR   SMART; SM00493; TOPRIM; 1.
DR   TIGRFAMs; TIGR01051; topA_bact; 1.
DR   PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW   DNA-binding; Isomerase; Topoisomerase; Complete proteome.
SQ   SEQUENCE   899 AA;  97723 MW;  1485DC4EDA0DA6FA CRC64;

  Query Match             41.3%;  Score 88;  DB 16;  Length 899;
  Best Local Similarity   58.7%;  Pred. No. 0.71;
  Matches   27;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYAKKAKAEKAK-KAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| | |:| ||   | |||.|  :| ||:| |||:|| :|
Db        836 AKKPAKKAAATKSKAKAESDAPAKKTAA-KKPAAKKPAAKKAAPKA 880


RESULT 9
Q8XVN7
ID   Q8XVN7      PRELIMINARY;      PRT;   200 AA.
AC   Q8XVN7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable histone H1 protein.
GN   RSC2793 OR RS00453.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Ralstoniaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646071; CAD16500.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
KW   Complete proteome.
SQ   SEQUENCE   200 AA;  19279 MW;  D3831B590510272D CRC64;

  Query Match             40.6%;  Score 86.5;  DB 16;  Length 200;
  Best Local Similarity   56.5%;  Pred. No. 0.24;
  Matches   26;  Conservative    4;  Mismatches   11;  Indels    5;  Gaps    2;

Qy          1 AKKY-AKKAKAEKAKKAYKAA----EAKKAAKYEKAAAEKAAAKEA 41
              |||  |||  |:||  | |||     |||||  :||| :| |||:|
Db         42 AKKVAAKKVAAKKAPAAKKAAVKKVAAKKAAPAKKAAVKKVAAKKA 87


RESULT 10
Q8PI40
ID   Q8PI40      PRELIMINARY;      PRT;   155 AA.
AC   Q8PI40;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Histone H1.
GN   XAC3058.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=92829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
DR   EMBL; AE011948; AAM37903.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
KW   Complete proteome.
SQ   SEQUENCE   155 AA;  16124 MW;  FDB53ECEB15F4D48 CRC64;

  Query Match             40.4%;  Score 86;  DB 16;  Length 155;
  Best Local Similarity   52.0%;  Pred. No. 0.21;
```

```
  Matches   26;  Conservative    4;  Mismatches   14;  Indels    6;  Gaps    2;

Qy          2 KKYAKKAKAEK-----AKKAYKAAEAKKAA-KYEKAAAEKAAAKEAAYEA 45
              || |||| |:|     |||| |   ||||  |  | | :| |||: | :|
Db         49 KKTAKKATAKKAVKKTAKKATKKTAAKKAVKKTAKKATKKTAAKKTAKKA 98


RESULT 11
Q8FP30
ID   Q8FP30      PRELIMINARY;      PRT;   168 AA.
AC   Q8FP30;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative 30S ribosomal protein S16.
GN   CE1960.
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA   Usuda Y., Sugimoto S.;
RT   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005220; BAC18770.1; -.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   168 AA;  18228 MW;  443CE68AA2321C01 CRC64;

  Query Match             40.4%;  Score 86;  DB 16;  Length 168;
  Best Local Similarity   65.7%;  Pred. No. 0.22;
  Matches   23;  Conservative    3;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          7 KAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              :|  || ||| :  |||:||  |||||||||| ||
Db        120 EAITEKKKKAREEKEAKEAA--EKAAAEKAAAAEA 152


RESULT 12
O93946
ID   O93946      PRELIMINARY;      PRT;   212 AA.
AC   O93946;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CTA2p (Fragment).
GN   CTA2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kaiser B., Kunkel W., Saluz H.P., Munder T.;
RT   "Identification of Candida albicans protein domains with
RT   transcriptional activating properties in Saccharomyces cerevisiae.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ006637; CAA07165.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   212 AA;  24231 MW;  10C2122E9554A387 CRC64;

  Query Match             40.4%;  Score 86;  DB 3;  Length 212;
  Best Local Similarity   48.9%;  Pred. No. 0.28;
  Matches   23;  Conservative   10;  Mismatches   12;  Indels    2;  Gaps    1;

Qy          1 AKKYAKKA--KAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |::|  | |:|||| :| :|::| | |:|| :   ||:|  ||
Db         70 AKKKAEEAKKKEEEAKKAEEAKKAEEAKKVEEAAKKAEEAKKAEEEA 116


RESULT 13
Q8WQ44
ID   Q8WQ44      PRELIMINARY;      PRT;    98 AA.
AC   Q8WQ44;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   18 kDa nuclear protein.
GN   LNP18.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tzortzakis N., Papageorgiou F.T., Tzinia A.K., Soteriadou K.P.;
RT   "Identification and characterization of a novel Leishmania gene
RT   encoding for a putative histone H1- like transcription factor.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ237814; CAD21431.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
KW   Nuclear protein.
SQ   SEQUENCE   98 AA;  9999 MW;  0A4AB93089D6C261 CRC64;

  Query Match             39.9%;  Score 85;  DB 5;  Length 98;
  Best Local Similarity   55.0%;  Pred. No. 0.17;
  Matches   22;  Conservative    3;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKE 40
              ||| |||   : |||| |    | | |  | ||:|||||:
Db         59 AKKPAKKVAKKPAKKAAKKPAKKPAKKAAKKAAKKAAAKK 98


RESULT 14
Q8T9R3
ID   Q8T9R3      PRELIMINARY;      PRT;   111 AA.
AC   Q8T9R3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   18 kDa nuclear protein.
OS   Leishmania infantum.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5671;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Papageorgiou F., Soteriadou K.;
RT   "Identification of a Leishmania infantum gene encoding for an histone
RT   H1-like nuclear protein.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF469106; AAL76335.1; -.
KW   Nuclear protein.
SQ   SEQUENCE   111 AA;  11162 MW;  16168F3B54960E83 CRC64;

  Query Match             39.9%;  Score 85;  DB 5;  Length 111;
  Best Local Similarity   55.0%;  Pred. No. 0.19;
  Matches   22;  Conservative    3;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKE 40
              ||| |||   : |||| |    | | |  | ||:|||||:
Db         72 AKKPAKKVAKKPAKKAAKKPAKKPAKKAAKKAAKKAAAKK 111


RESULT 15
Q45370
ID   Q45370      PRELIMINARY;      PRT;   182 AA.
AC   Q45370;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Histone H1.
GN   H1.
OS   Bordetella pertussis.
```

```
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BPW28;
RX   MEDLINE=95319329; PubMed=7596289;
RA   Scarlato V., Arico B., Goyard S., Ricci S., Manetti R., Prugnola A.,
RA   Manetti R., Polverino-De-Laureto P., Ullmann A., Rappuoli R.;
RT   "A novel chromatin-forming histone H1 homologue is encoded by a
RT   dispensable and growth-regulated gene in Bordetella pertussis.";
RL   Mol. Microbiol. 15:871-881(1995).
DR   EMBL; L37438; AAB59120.1; -.
SQ   SEQUENCE   182 AA;  18252 MW;  9A17A397B12B0421 CRC64;

  Query Match             39.9%;  Score 85;  DB 2;  Length 182;
  Best Local Similarity   61.4%;  Pred. No. 0.3;
  Matches   27;  Conservative    4;  Mismatches    9;  Indels    4;  Gaps    3;

Qy          1 AKKYAKKAKAEK--AKKAY-KAAEAKKAAKYEKAAAEKAAAKEA 41
              ||| |||| |:|  ||||  | | ||||   :|| |:|| ||:|
Db         58 AKKVAKKAVAKKAVAKKAVAKKAVAKKAVA-KKAVAKKAVAKKA 100


Search completed: January 21, 2004, 09:00:26
Job time : 21.1169 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:52:56 ; Search time 21.6234 Seconds
(without alignments)
330.323 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	213	100.0	45	21	AAY82572	Copolymer molecula
2	138	64.8	109	21	AAY82577	Copolymer molecula
3	134.5	63.1	56	21	AAY82573	Copolymer molecula
4	131	61.5	77	21	AAY82575	Copolymer molecula
5	126.5	59.4	86	21	AAY82576	Copolymer molecula
6	120.5	56.6	66	21	AAY82574	Copolymer molecula
7	108	50.7	35	21	AAY82571	Copolymer molecula
8	99.5	46.7	106	11	AAR06446	Recombinant copoly
9	98	46.0	46	13	AAR28871	High affinity macr
10	94.5	44.4	239	23	ABB49123	Listeria monocytog
11	94	44.1	154	11	AAR06445	Recombinant copoly
12	91	42.7	223	20	AAY14928	Amino acid sequenc
13	89.5	42.0	214	20	AAY34055	M. tuberculosis hi
14	89.5	42.0	214	21	AAY57353	M. tuberculosis hi
15	87	40.8	334	22	ABG28693	Novel human diagno
16	86	40.4	39	19	AAW44934	Mycobacterial hepa
17	86	40.4	198	19	AAW43082	Mycobacterial hepa
18	86	40.4	198	19	AAW44936	Mycobacterial hepa
19	86	40.4	347	24	ABJ18771	Pseudomonas aerugi
20	85.5	40.1	205	21	AAB20575	Mycobacterium bovi
21	85	39.9	427	22	AAG70868	C albicans apoptos
22	82	38.5	165	22	AAG91997	C glutamicum prote
23	81	38.0	80	23	ABG71044	Tumour neoangiogen
24	77.5	36.4	299	22	ABB63276	Drosophila melanog
25	77.5	36.4	2151	22	ABB60086	Drosophila melanog
26	77	36.2	103	20	AAY34067	Histone H1 isoform
27	77	36.2	103	21	AAY57365	Human histone H1.5
28	77	36.2	116	20	AAY34066	Histone H1 isoform
29	77	36.2	116	21	AAY57364	Human histone H1.5
30	77	36.2	158	20	AAY34068	Histone H1 isoform
31	77	36.2	158	21	AAY57366	Human histone H1.5
32	77	36.2	222	18	AAW29476	Human histone H1 i
33	77	36.2	222	20	AAY34033	Human histone H1 i
34	77	36.2	222	21	AAY57331	Human histone H1 i
35	77	36.2	226	20	AAY34060	Human histone H1 i
36	77	36.2	226	21	AAY57358	Human histone H1 i
37	76	35.7	157	22	ABB58855	Drosophila melanog
38	75.5	35.4	111	23	ABP00299	Human ORFX protein
39	75.5	35.4	201	23	ABP41465	Human ovarian anti
40	75.5	35.4	216	23	ABP62982	Human polypeptide
41	75.5	35.4	218	21	AAG00755	Human secreted pro
42	75.5	35.4	218	21	AAG00759	Human secreted pro
43	75.5	35.4	220	23	AAU76972	Human ribosomal L1
44	75.5	35.4	220	24	ABU07443	Protein differenti
45	75.5	35.4	221	23	ABP62855	Human polypeptide

ALIGNMENTS

RESULT 1
AAY82572
ID AAY82572 standard; peptide; 45 AA.
XX
AC AAY82572;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:2.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 45 AA;

Query Match 100.0%; Score 213; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
||||||||||||||||||||||||||||||||||||||||||||
Db 1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45

RESULT 2
AAY82577
ID AAY82577 standard; peptide; 109 AA.
XX
AC AAY82577;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 109 AA;

Query Match 64.8%; Score 138; DB 21; Length 109;
Best Local Similarity 76.6%; Pred. No. 5.9e-08;
Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

Qy 1 AKKYAKKAKAEKAKKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
|||||| |||| || | |||||| ||| || | ||||||||||
Db 65 AKKYAKAAKAE--KKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109

RESULT 3
AAY82573
ID AAY82573 standard; peptide; 56 AA.
XX
AC AAY82573;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.

XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 56 AA;

Query Match 63.1%; Score 134.5; DB 21; Length 56;
Best Local Similarity 68.4%; Pred. No. 7.2e-08;
Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

```
Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
              |||||||      |||||      ||||||||||   ||| || | ||||||||||||
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAEAKY-KAEAAKAAAKEAAYEA 56
```

RESULT 4
AAY82575
ID AAY82575 standard; peptide; 77 AA.
XX
AC AAY82575;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 77 AA;

Query Match 61.5%; Score 131; DB 21; Length 77;
Best Local Similarity 50.6%; Pred. No. 2.4e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

```
Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||      |||||      |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 -KAAAEKAAAKEAAYEA 45
               || | ||||||||||
Db         61 YKAEAAKAAAKEAAYEA 77
```

RESULT 5
AAY82576
ID AAY82576 standard; peptide; 86 AA.
XX
AC AAY82576;
XX
DT 28-JUL-2000 (first entry)
XX

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 86 AA;

Query Match 59.4%; Score 126.5; DB 21; Length 86;
Best Local Similarity 45.3%; Pred. No. 8.3e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3;

```
Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||      |||||      |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 ----------KAAAEKAAAKEAAYEA 45
                      || | ||||||||||||
Db         61 YKAEAAKKAYKAEAAKAAAKEAAYEA 86
```

RESULT 6
AAY82574
ID AAY82574 standard; peptide; 66 AA.
XX
AC AAY82574;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 66 AA;

Query Match 56.6%; Score 120.5; DB 21; Length 66;
Best Local Similarity 71.1%; Pred. No. 2.8e-06;
Matches 32; Conservative 0; Mismatches 6; Indels 7; Gaps 2;

```
Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |||||| ||||  || | |||||      || | ||||||||||
Db         29 AKKYAKAAKAE--KKEYAAAEAK-----YKAEAAKAAAKEAAYEA 66
```

RESULT 7
AAY82571
ID AAY82571 standard; peptide; 35 AA.
XX
AC AAY82571;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:1.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 35 AA;

```
  Query Match             50.7%;  Score 108;  DB 21;  Length 35;
  Best Local Similarity   64.4%;  Pred. No. 3.4e-05;
  Matches   29;  Conservative    1;  Mismatches    5;  Indels   10;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||||||| ||  ||||||         |  || | :||||||||||
Db          1 AKKYAKKEKA--AKKAYK--------KEAKAKAAEAAAKEAAYEA 35
```

RESULT 8
AAR06446
ID AAR06446 standard; protein; 106 AA.
XX
AC AAR06446;
XX
DT 25-MAR-2003 (updated)
DT 03-JAN-1991 (first entry)
XX
DE Recombinant copolymer 1-19, myelin basic protein analogue.
XX
KW Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;
KW immunological activity; autoimmune encephalomyelitis;
KW multiple sclerosis;
XX
OS Synthetic.
XX
PN EP383620-A.
XX
PD 22-AUG-1990.
XX
PF 16-FEB-1990; 90EP-0301700.
XX
PR 07-FEB-1990; 90US-0473845.
PR 17-FEB-1989; 89US-0312541.
XX
PA (REPK) REPLIGEN CORP.
XX
PI Cook KS;
XX
DR WPI; 1990-255848/34.
DR N-PSDB; AAQ06446.
XX
PT Producing genes encoding random polymers of aminoacid(s) - for
PT producing recombinant polypeptide(s) with biological and/or
PT immunological activity
XX
PS Disclosure; Fig 12; 25pp; English.
XX
CC To improve the expression of rCOP-1 polypeptides in E. coli, genes
CC coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN
CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to
CC express Protein A. The resulting plasmids encode fusion proteins
CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.
CC A methionine residue occurs between the Protein A and rCOP-1
CC sequences, originating from the 5' linker sequence, in order that
CC the COP-1 polypeptide may be cleaved from the fusion protein.
CC rCOP-1-19 contains oligonucleotide duplexes incoding the following
CC segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The
CC N-terminal alanine residue is left behind following CNBr cleavage of the
CC fusion protein.
CC The product prevents or arrests experimental autoimmune
CC encephalomyelitis. They are used to prevent, arrest or control a
CC demyelinating disorder, e.g. multiple sclerosis. They may also
CC be used as additives to hair care products to confer beneficial
CC effects on damaged hair or as supplements for diets deficient in
CC certain amino acids.
CC See also AAQ05664.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 106 AA;

```
  Query Match             46.7%;  Score 99.5;  DB 11;  Length 106;
  Best Local Similarity   58.3%;  Pred. No. 0.00089;
  Matches   28;  Conservative    4;  Mismatches    7;  Indels    9;  Gaps    2;

Qy          3 KYAKKA-KAEKAKKAYKAAEAKKAAKYEKAAAEKAA--------AKEA 41
              | |||| :|||||  |:| :|:|| | ||||||| |        ||||
Db          8 KAAKKAYEAEKAKAKYEAKKAEKAEKAEKAAAEKKAKEAKKAEKAKEA 55


RESULT 9
AAR28871
ID   AAR28871 standard; peptide; 46 AA.
XX
AC   AAR28871;
XX
DT   25-MAR-2003  (updated)
DT   23-MAR-1993  (first entry)
XX
DE   High affinity macrophage mannose receptor ligand compound #9.
XX
KW   glycopeptide; mannose; mannosylated; glycosylated; mannose receptor;
KW   macrophages; monocytes; destroy; cytotoxicity; label; image; alter;
KW   macrophage processing of antigen; MHC restriction; inflammation;
KW   inflammatory diseases; macrophage secretory products;
KW   Crohn's disease; legionnaires disease; mononuclear phagocytes; HIV;
KW   AIDS; lysosomal storage diseases; Gaucher's disease; asthma;
KW   alveolar macrophages metastasis; systemic macrophages; deliver;
KW   antigenic peptides; prevent transplant rejection;
KW   organ transplantation; antitumour agents; cancer; toxins.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine. May also have non
FT                         interfering substits."
FT   Modified-site   4
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   7
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   10
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   13
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   16
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   19
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   22
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   25
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   28
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   31
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   34
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   37
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   40
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   43
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine."
FT   Modified-site   46
FT                   /note= "opt may have mannose, fucose, glucose or
FT                         N-Ac-glucosamine. May also have non
FT                         interfering substits."
XX
PN   WO9219248-A1.
XX
PD   12-NOV-1992.
XX
PF   01-MAY-1992;   92WO-US03609.
XX
PR   03-MAY-1991;   91US-0694983.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Stahl PD;
XX
DR   WPI; 1992-398516/48.
XX
PT   New high affinity mannose receptor ligand cpds. - for treating
PT   diseases mediated by macrophage activity e.g. asthma,
PT   inflammatory diseases and infectious diseases, e.g. HIV
XX
PS   Claim 3; Page 21; 32pp; English.
XX
CC   This compound represents a glycopeptide effective in inhibiting the
CC   binding of labelled mannosylated BSA to mannose receptors. Mannose
CC   receptors are uniquely found on macrophages and not on monocytes.
CC   Glycopeptides such as this provide a mechanism to target macrophages
CC   specifically, to image, label, destroy or otherwise alter their
CC   antigen processing function. In addition they can be conjugated to
CC   solid supports and used to purify mannose receptors from a variety
CC   of sources. They are useful in the treatment of inflammatory
CC   diseases driven by macrophage secretory products eg. Crohn's
CC   disease; infectious diseases in which macrophages harbour replicating
CC   infectious agents eg. Legionnaires disease; viral infections
CC   involving mononuclear phagocytes eg. HIV and lysosomal storage
CC   diseases, in which macrophages are principally involved eg.
CC   Gaucher's disease; asthma mediated by alveolar macrophages; and in
CC   controlling metastasis, mediated by systemic macrophages. The peptides
CC   can also be used to deliver antigenic peptides as conjugates to a
CC   macrophage to marshal an immune response; also self peptides to
CC   prevent tissue transplant rejection.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   46 AA;

  Query Match             46.0%;  Score 98;  DB 13;  Length 46;
  Best Local Similarity   63.4%;  Pred. No. 0.00055;
  Matches   26;  Conservative    1;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              | | || ||| || || |||:| ||||  |||    ||| |
Db          2 AAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 42


RESULT 10
ABB49123
ID   ABB49123 standard; Protein; 239 AA.
XX
AC   ABB49123;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes protein #1827.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
```

```
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI   Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI   Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI   Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI   Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI   Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI   Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI   Rose M, Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides -
XX
PS   Claim 6; SEQ ID No 1828; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   239 AA;

  Query Match             44.4%;  Score 94.5;  DB 23;  Length 239;
  Best Local Similarity   56.2%;  Pred. No. 0.0072;
  Matches   27;  Conservative    7;  Mismatches    9;  Indels    5;  Gaps    2;

Qy          2 KKYAKKAKAEKAKK---AYKAAEAKK--AAKYEKAAAEKAAAKEAAYE 44
              |  |:||:|:| |:   | ||| |||   |  |||||:||||::|| |
Db        124 KAAAEKAEADKKKQEEDAVKAANAKKEQEAAEEKAAADKAAAEKAAAE 171


RESULT 11
AAR06445
ID   AAR06445 standard; protein; 154 AA.
XX
AC   AAR06445;
XX
DT   25-MAR-2003  (updated)
DT   03-JAN-1991  (first entry)
XX
DE   Recombinant copolymer 1-77, myelin basic protein analogue.
XX
KW   Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;
KW   immunological activity; autoimmune encephalomyelitis;
KW   multiple sclerosis;
XX
OS   Synthetic.
XX
PN   EP383620-A.
XX
PD   22-AUG-1990.
XX
PF   16-FEB-1990;   90EP-0301700.
XX
PR   07-FEB-1990;   90US-0473845.
PR   17-FEB-1989;   89US-0312541.
XX
PA   (REPK ) REPLIGEN CORP.
XX
PI   Cook KS;
XX
DR   WPI; 1990-255848/34.
DR   N-PSDB; AAQ05664.
XX
PT   Producing genes encoding random polymers of aminoacid(s) - for
PT   producing recombinant polypeptide(s) with biological and/or
PT   immunological activity
XX
PS   Disclosure; Fig 11; 25pp; English.
XX
CC   To improve the expression of rCOP-1 polypeptides in E. coli, genes
CC   coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN
CC   (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to
CC   express Protein A. The resulting plasmids encode fusion proteins
CC   consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.
CC   A methionine residue occurs between the Protein A and rCOP-1
CC   sequences, originating from the 5' linker sequence, in order that
CC   the COP-1 polypeptide may be cleaved from the fusion protein.
CC   rCOP-1-77 contains oligonucleotide duplexes incoding the following
CC   segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue
CC   is left behind following CNBr cleavage of the fusion protein.
CC   The product prevents or arrests experimental autoimmune
CC   encephalomyelitis. They are used to prevent, arrest or control a
CC   demyelinating disorder, e.g. multiple sclerosis. They may also
CC   be used as additives to hair care products to confer beneficial
CC   effects on damaged hair or as supplements for diets deficient in
CC   certain amino acids.
CC   See also AAQ05665.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   154 AA;

  Query Match             44.1%;  Score 94;  DB 11;  Length 154;
  Best Local Similarity   60.5%;  Pred. No. 0.0052;
  Matches   26;  Conservative    2;  Mismatches    9;  Indels    6;  Gaps    2;

Qy          2 KKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYE 44
              ||| ||||  | ||  || ||:||    |||||   |||| |:
Db        105 KKYKKKAKKAKYKK--KAKEAEKA----KAAAEAEKAKEAEYK 141


RESULT 12
AAY14928
ID   AAY14928 standard; protein; 223 AA.
XX
AC   AAY14928;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Amino acid sequence of M. vaccae antigen GV-45.
XX
KW   Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW   dendritic cell maturation; infectious disease; immune disorder; cancer;
KW   respiratory system; mycobacterial infection; allergy; tuberculosis;
KW   leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW   dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
```

KW squamous cell carcinoma; melanoma.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI; 1999-430163/36.
DR N-PSDB; AAZ11393.
XX
PT Enhancing immune response to an antigen
XX
PS Claim 1; Page 239; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 223 AA;

Query Match 42.7%; Score 91; DB 20; Length 223;
Best Local Similarity 59.6%; Pred. No. 0.016;
Matches 28; Conservative 5; Mismatches 12; Indels 2; Gaps 2;

```
Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKAAKYEKAA-AEKAAAKEAAYEA 45
              ||| | | | |:||  | ||| ||||   :||| |:|| ||:|| :|
Db        137 AKKAATKAAPAKKATAAKKAAPAKKATAAKKAAPAKKAPAKKAATKA 183
```

RESULT 13
AAY34055
ID AAY34055 standard; protein; 214 AA.
XX
AC AAY34055;
XX
DT 23-NOV-1999 (first entry)
XX
DE M. tuberculosis histone H1-like antigen.
XX
KW Ulcerative colitis; histone; H1-like antigen; porin antigen;
KW Bacteroides antigen; inflammatory bowel disease; IBD; pANCA; diagnosis;
KW perinuclear anti-neutrophil cytoplasmic antibody.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9945955-A1.
XX
PD 16-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05492.
XX
PR 12-MAR-1998; 98US-0041889.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Braun J, Cohavy O;
XX
DR WPI; 1999-551215/46.
XX
PT Use of histone H1, porin or Bacteroides antigens as targets for the
PT diagnosis, prevention and treatment of ulcerative colitis -
XX
PS Claim 2; Fig 11; 134pp; English.
XX
CC The invention provides a method for the diagnosis, prevention and
CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a
CC porin antigen or a Bacteroides antigen as a target antigen. The novel
CC method of diagnosing UC in a subject suspected of having inflammatory
CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
CC (2) contacting the sample with a histone H1-like antigen, or perinuclear
CC anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
CC complex of the histone H1-like antigen, or the pANCA-reactive fragment,
CC and antibody to the histone H1-like antigen; and (3) detecting the
CC presence or absence of the complex; where the presence of the complex
CC indicates that the subject has UC. The pANCA-reactive histone H1-like
CC antigen, porin antigen and Bacteroides antigen are useful in the
CC diagnosis, prevention and treatment of UC. The methods can also be used
CC for identifying agents useful for treating UC. The present sequence
CC represents a M. tuberculosis histone H1-like antigen.
XX
SQ Sequence 214 AA;

Query Match 42.0%; Score 89.5; DB 20; Length 214;
Best Local Similarity 55.6%; Pred. No. 0.023;
Matches 25; Conservative 4; Mismatches 13; Indels 3; Gaps 1;

```
Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  |  |   : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152
```

RESULT 14
AAY57353
ID AAY57353 standard; Protein; 214 AA.
XX
AC AAY57353;
XX
DT 13-JUN-2000 (first entry)
XX
DE M. tuberculosis histone H1-like protein, 214.
XX
KW Ulcerative colitis; inflammatory bowel disease; porin antigen; MAb;
KW pANCA; perinuclear anti-neutrophil cytoplasmic antibody; 214 protein;
KW histone H1; isoform.
XX
OS Mycobacterium tuberculosis.
XX
PN US6033864-A.
XX
PD 07-MAR-2000.
XX
PF 12-MAR-1998; 98US-0041889.
XX
PR 12-APR-1996; 96US-0057846.
PR 11-APR-1997; 97US-0837058.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Cohavy O, Braun J;

```
XX
DR   WPI; 2000-255695/22.
XX
PT   Diagnosing ulcerative colitis or susceptibility, by detecting complex
PT   formation between microbial porin antigen and perinuclear
PT   anti-neutrophil cytoplasmic autoantibodies -
XX
PS   Example 5; Fig 10; 49pp; English.
XX
CC   The invention provides a method for diagnosing ulcerative colitis in a
CC   subject suspected of having inflammatory bowel disease. The method
CC   comprises reacting a patient sample with a porin antigen that is
CC   immunologically reactive with pANCA (perinuclear anti-neutrophil
CC   cytoplasmic antibodies) and detecting formation of a Ag-pANCA complex
CC   as indicative of ulcerative colitis. The method is used to diagnose
CC   ulcerative colitis or susceptibility to it. The present sequence
CC   represents a histone H1-like protein of M. tuberculosis, designated 214.
XX
SQ   Sequence   214 AA;

  Query Match             42.0%;  Score 89.5;  DB 21;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.023;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  | |   : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 15
ABG28693
ID   ABG28693 standard; Protein; 334 AA.
XX
AC   ABG28693;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #28684.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92880.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 59052; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   334 AA;

  Query Match             40.8%;  Score 87;  DB 22;  Length 334;
  Best Local Similarity   57.4%;  Pred. No. 0.067;
  Matches   27;  Conservative    6;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYAKKAKAEKAK-KAYKAAEAKKAAKYEKAAAEKAAA-KEAAYEA 45
              |:| |    ||||| :| | | |:|||  :||||||||  |:|| :|
Db        214 ARKKAATETAEKAKAQADKKAAAEKAAADKKAAAEKAATDKKAAEKA 260


Search completed: January 21, 2004, 08:59:09
Job time : 22.6234 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:01:02 ; Search time 16.3636 Seconds
(without alignments)
562.323 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	213	100.0	45	10	US-09-816-989A-2	Sequence 2, Appli
2	138	64.8	109	10	US-09-816-989A-7	Sequence 7, Appli
3	134.5	63.1	56	10	US-09-816-989A-3	Sequence 3, Appli
4	131	61.5	77	10	US-09-816-989A-5	Sequence 5, Appli
5	126.5	59.4	86	10	US-09-816-989A-6	Sequence 6, Appli
6	120.5	56.6	66	10	US-09-816-989A-4	Sequence 4, Appli
7	108	50.7	35	10	US-09-816-989A-1	Sequence 1, Appli
8	91	42.7	223	12	US-10-205-979-52	Sequence 52, Appl
9	91	42.7	223	14	US-10-051-643-201	Sequence 201, App
10	89.5	42.0	214	15	US-10-229-567-27	Sequence 27, Appl
11	86	40.4	347	15	US-10-127-032-120	Sequence 120, App
12	85	39.9	376	15	US-10-156-761-9889	Sequence 9889, Ap
13	82	38.5	165	10	US-09-738-626-5751	Sequence 5751, Ap
14	78	36.6	272	15	US-10-156-761-12370	Sequence 12370, A
15	78	36.6	309	11	US-09-820-843A-24	Sequence 24, Appl
16	77.5	36.4	227	15	US-10-156-761-10923	Sequence 10923, A
17	77	36.2	103	15	US-10-229-567-39	Sequence 39, Appl
18	77	36.2	116	15	US-10-229-567-38	Sequence 38, Appl
19	77	36.2	158	15	US-10-229-567-40	Sequence 40, Appl
20	77	36.2	222	15	US-10-229-567-3	Sequence 3, Appli
21	77	36.2	226	15	US-10-229-567-32	Sequence 32, Appl
22	75.5	35.4	201	12	US-10-264-049-2597	Sequence 2597, Ap
23	75.5	35.4	220	9	US-09-923-304-2	Sequence 2, Appli
24	75.5	35.4	265	9	US-09-925-302-559	Sequence 559, App
25	74.5	35.0	1002	12	US-10-369-493-1378	Sequence 1378, Ap
26	74	34.7	352	11	US-09-820-843A-23	Sequence 23, Appl
27	74	34.7	356	11	US-09-820-843A-27	Sequence 27, Appl
28	74	34.7	942	15	US-10-156-761-12155	Sequence 12155, A
29	73	34.3	234	15	US-10-262-209-1	Sequence 1, Appli
30	72.5	34.0	372	11	US-09-820-843A-8	Sequence 8, Appli
31	71	33.3	130	15	US-10-262-209-2	Sequence 2, Appli
32	71	33.3	218	15	US-10-229-567-4	Sequence 4, Appli
33	71	33.3	228	15	US-10-156-761-9425	Sequence 9425, Ap
34	71	33.3	539	12	US-10-369-493-17058	Sequence 17058, A
35	70.5	33.1	548	15	US-10-128-714-3127	Sequence 3127, Ap
36	70.5	33.1	555	15	US-10-128-714-8127	Sequence 8127, Ap
37	70.5	33.1	619	11	US-09-882-774-1	Sequence 1, Appli
38	70	32.9	55	15	US-10-229-567-41	Sequence 41, Appl
39	70	32.9	347	15	US-10-156-761-11652	Sequence 11652, A
40	70	32.9	641	9	US-09-765-272-160	Sequence 160, App
41	70	32.9	744	11	US-09-769-787-184	Sequence 184, App
42	70	32.9	1156	9	US-09-815-242-13187	Sequence 13187, A
43	69.5	32.6	578	12	US-10-032-585-7390	Sequence 7390, Ap
44	69.5	32.6	1130	12	US-10-369-493-6751	Sequence 6751, Ap
45	69.5	32.6	1387	15	US-10-156-761-13000	Sequence 13000, A

ALIGNMENTS

RESULT 1
US-09-816-989A-2
; Sequence 2, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-2

Query Match 100.0%; Score 213; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
```

RESULT 2
US-09-816-989A-7
; Sequence 7, Application US/09816989A

```
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 109
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

  Query Match             64.8%;  Score 138;  DB 10;  Length 109;
  Best Local Similarity   76.6%;  Pred. No. 4.8e-08;
  Matches   36;  Conservative    0;  Mismatches    7;  Indels    4;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
              |||||| ||||  || | ||||||  |||  || | ||||||||||||
Db         65 AKKYAKAAKAE--KKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109


RESULT 3
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

  Query Match             63.1%;  Score 134.5;  DB 10;  Length 56;
  Best Local Similarity   68.4%;  Pred. No. 5.6e-08;
  Matches   39;  Conservative    0;  Mismatches    5;  Indels   13;  Gaps    4;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
              |||||||     |||||     ||||||||||   ||| || | ||||||||||||
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAEAKY-KAEAAKAAAKEAAYEA 56


RESULT 4
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 77
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

  Query Match             61.5%;  Score 131;  DB 10;  Length 77;
  Best Local Similarity   50.6%;  Pred. No. 1.9e-07;
  Matches   39;  Conservative    0;  Mismatches    6;  Indels   32;  Gaps    3;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||     |||||     |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 -KAAAEKAAAKEAAYEA 45
               || | ||||||||||||
Db         61 YKAEAAKAAAKEAAYEA 77


RESULT 5
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 86
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

  Query Match             59.4%;  Score 126.5;  DB 10;  Length 86;
  Best Local Similarity   45.3%;  Pred. No. 6.5e-07;
  Matches   39;  Conservative    0;  Mismatches    6;  Indels   41;  Gaps    3;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||     |||||     |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 ----------KAAAEKAAAKEAAYEA 45
                        || | ||||||||||||
Db         61 YKAEAAKKAYKAEAAKAAAKEAAYEA 86


RESULT 6
```

```
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

  Query Match             56.6%;  Score 120.5;  DB 10;  Length 66;
  Best Local Similarity   71.1%;  Pred. No. 2.2e-06;
  Matches   32;  Conservative    0;  Mismatches    6;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |||||| ||||  || | |||||      || | ||||||||||||
Db         29 AKKYAKAAKAE--KKEYAAAEAK-----YKAEAAKAAAKEAAYEA 66


RESULT 7
US-09-816-989A-1
; Sequence 1, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 35
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-1

  Query Match             50.7%;  Score 108;  DB 10;  Length 35;
  Best Local Similarity   64.4%;  Pred. No. 2.5e-05;
  Matches   29;  Conservative    1;  Mismatches    5;  Indels   10;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||||||| ||  ||||||         |  || | :||||||||||
Db          1 AKKYAKKEKA--AKKAYK--------KEAKAKAAEAAAKEAAYEA 35


RESULT 8
US-10-205-979-52
; Sequence 52, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Tan, Paul L. J.
;  APPLICANT:  Abernethy, Nevin
;  TITLE OF INVENTION: Compounds and Methods for the Modulation
;  TITLE OF INVENTION:  of Immune Responses
;  FILE REFERENCE: 11000.1063U
;  CURRENT APPLICATION NUMBER: US/10/205,979
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/308,446
;  PRIOR FILING DATE: 2001-07-26
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-10-205-979-52

  Query Match             42.7%;  Score 91;  DB 12;  Length 223;
  Best Local Similarity   59.6%;  Pred. No. 0.013;
  Matches   28;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKAAKYEKAA-AEKAAAKEAAYEA 45
              ||| | | | |:||  | ||| ||||   :||| |:|| ||:|| :|
Db        137 AKKAATKAAPAKKATAAKKAAPAKKATAAKKAAPAKKAPAKKAATKA 183


RESULT 9
US-10-051-643-201
; Sequence 201, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Tan, Paul L. J.
;  TITLE OF INVENTION: Methods and Compounds for the Treatment
;  TITLE OF INVENTION:  of Immunologically-Mediated Diseases of the Respiratory
;  TITLE OF INVENTION:  System using Mycobacterium Vaccae
;  FILE REFERENCE: 11000.1008c2
;  CURRENT APPLICATION NUMBER: US/10/051,643
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US09/156,181
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: US 08/996,624
;  PRIOR FILING DATE: 1997-12-23
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 201
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-10-051-643-201

  Query Match             42.7%;  Score 91;  DB 14;  Length 223;
  Best Local Similarity   59.6%;  Pred. No. 0.013;
  Matches   28;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKAAKYEKAA-AEKAAAKEAAYEA 45
              ||| | | | |:||  | ||| ||||   :||| |:|| ||:|| :|
Db        137 AKKAATKAAPAKKATAAKKAAPAKKATAAKKAAPAKKAPAKKAATKA 183


RESULT 10
US-10-229-567-27
; Sequence 27, Application US/10229567
; Publication No. US20030092080A1
;   GENERAL INFORMATION:
;        APPLICANT: Braun, Jonathan
;                   Cohavy, Offer
;        TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;                            Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;                            Microbial UC pANCA antigens
;        NUMBER OF SEQUENCES: 41
```

```
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Campbell & Flores LLP
;             STREET: 4370 La Jolla Village Drive, Suite 700
;             CITY: San Diego
;             STATE: California
;             COUNTRY: USA
;             ZIP: 92122
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/229,567
;             FILING DATE: 27-Aug-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/417,264
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: US 09/041,889
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Campbell, Cathryn A.
;             REGISTRATION NUMBER: 31,815
;             REFERENCE/DOCKET NUMBER: P-PM 3006
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (619) 535-9001
;             TELEFAX: (619) 535-8949
;    INFORMATION FOR SEQ ID NO: 27:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 214 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-10-229-567-27

  Query Match             42.0%;  Score 89.5;  DB 15;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.018;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  |  |   : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 11
US-10-127-032-120
; Sequence 120, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
;  APPLICANT: Whiteley, Marvin
;  APPLICANT:  Bangera, M. Gita
;  APPLICANT:  Lory, Stephen
;  APPLICANT:  Greenberg, Everett Peter
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
;  TITLE OF INVENTION:  BIOFILM FORMATION
;  FILE REFERENCE: UIZ-070CP
;  CURRENT APPLICATION NUMBER: US/10/127,032
;  CURRENT FILING DATE:  2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,190
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: US 60/344,142
;  PRIOR FILING DATE: 2001-10-24
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 120
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-10-127-032-120

  Query Match             40.4%;  Score 86;  DB 15;  Length 347;
  Best Local Similarity   51.1%;  Pred. No. 0.072;
  Matches   24;  Conservative    6;  Mismatches   15;  Indels    2;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKA--YKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |:| |:  ||::||||   |||| || |   |  ||  | |:|| :|
Db        125 AQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKRAEDEAKKKAAEDA 171


RESULT 12
US-10-156-761-9889
; Sequence 9889, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9889
;   LENGTH: 376
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9889

  Query Match             39.9%;  Score 85;  DB 15;  Length 376;
  Best Local Similarity   50.0%;  Pred. No. 0.1;
  Matches   21;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              ||| |   |:   | | | | ||| |  :|:||:|:|||: |
Db        310 AKKSAPAKKSAPGKTAAKKAAAKKTAPAKKSAAKKSAAKKTA 351


RESULT 13
US-09-738-626-5751
; Sequence 5751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5751
;   LENGTH: 165
```

```
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

  Query Match             38.5%;  Score 82;  DB 10;  Length 165;
  Best Local Similarity   62.9%;  Pred. No. 0.087;
  Matches   22;  Conservative    4;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          7 KAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              :|  || ||| :  |||:||  |||||||||| |:
Db        120 EAITEKKKKAREDKEAKEAA--EKAAAEKAAAAES 152


RESULT 14
US-10-156-761-12370
; Sequence 12370, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12370
;   LENGTH: 272
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-12370

  Query Match             36.6%;  Score 78;  DB 15;  Length 272;
  Best Local Similarity   46.3%;  Pred. No. 0.41;
  Matches   19;  Conservative    7;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEA 41
              | | ||:||:: |    ||   |||    :||||:||:: |
Db         90 AAKAAKQAKSDLADAKKKAEATKKAEAARRAAAERAASRSA 130


RESULT 15
US-09-820-843A-24
; Sequence 24, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
;  TITLE OF INVENTION:  USEFUL AS ANTI-INFECTIVES
;  FILE REFERENCE: Q63915
;  CURRENT APPLICATION NUMBER: US/09/820,843A
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
;   LENGTH: 309
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: polyhydroxyalkanoate synthesis protein PhaF
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: gi|9951352
US-09-820-843A-24

  Query Match             36.6%;  Score 78;  DB 11;  Length 309;
  Best Local Similarity   54.8%;  Pred. No. 0.47;
  Matches   23;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              ||  || |    || | | | || |||    || | ||| ||
Db        142 AKAAAKPAAKPAAKPAAKTAAAKPAAKPAAKAAAKPAAKPAA 183


Search completed: January 21, 2004, 09:10:08
Job time : 16.3636 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:57:47 ; Search time 8.47403 Seconds
(without alignments)
224.685 Million cell updates/sec

Title: US-09-816-989A-2
Perfect score: 213
Sequence: 1 AKKYAKKAKAEKAKKAYKAA.....AKYEKAAAEKAAAKEAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	213	100.0	45	4	US-09-405-743A-2	Sequence 2, Appli
2	138	64.8	109	4	US-09-405-743A-7	Sequence 7, Appli
3	134.5	63.1	56	4	US-09-405-743A-3	Sequence 3, Appli
4	131	61.5	77	4	US-09-405-743A-5	Sequence 5, Appli
5	126.5	59.4	86	4	US-09-405-743A-6	Sequence 6, Appli
6	120.5	56.6	66	4	US-09-405-743A-4	Sequence 4, Appli
7	108	50.7	35	4	US-09-405-743A-1	Sequence 1, Appli
8	91	42.7	223	3	US-09-095-855-201	Sequence 201, App
9	91	42.7	223	4	US-09-205-426-201	Sequence 201, App
10	89.5	42.0	214	3	US-09-041-889-27	Sequence 27, Appl
11	89.5	42.0	214	4	US-09-417-264-27	Sequence 27, Appl
12	86	40.4	407	4	US-09-252-991A-29581	Sequence 29581, A
13	78	36.6	316	4	US-09-252-991A-32957	Sequence 32957, A
14	77.5	36.4	700	4	US-09-107-532A-5094	Sequence 5094, Ap
15	77	36.2	103	3	US-09-041-889-39	Sequence 39, Appl
16	77	36.2	103	4	US-09-417-264-39	Sequence 39, Appl
17	77	36.2	116	3	US-09-041-889-38	Sequence 38, Appl
18	77	36.2	116	4	US-09-417-264-38	Sequence 38, Appl
19	77	36.2	158	3	US-09-041-889-40	Sequence 40, Appl
20	77	36.2	158	4	US-09-417-264-40	Sequence 40, Appl
21	77	36.2	222	3	US-09-041-889-3	Sequence 3, Appli
22	77	36.2	222	3	US-08-837-058-3	Sequence 3, Appli
23	77	36.2	222	4	US-09-417-264-3	Sequence 3, Appli
24	77	36.2	226	3	US-09-041-889-32	Sequence 32, Appl
25	77	36.2	226	4	US-09-417-264-32	Sequence 32, Appl
26	75	35.2	8991	4	US-08-714-741-32	Sequence 32, Appl
27	74.5	35.0	32	1	US-08-152-488-13	Sequence 13, Appl
28	74.5	35.0	32	1	US-08-303-025-15	Sequence 15, Appl
29	74.5	35.0	32	1	US-08-677-304-13	Sequence 13, Appl
30	74.5	35.0	32	2	US-08-436-703B-2	Sequence 2, Appli
31	74.5	35.0	33	1	US-08-303-025-16	Sequence 16, Appl
32	74.5	35.0	33	2	US-08-436-703B-4	Sequence 4, Appli
33	74	34.7	399	4	US-09-252-991A-22853	Sequence 22853, A
34	73.5	34.5	623	4	US-08-714-741-47	Sequence 47, Appl
35	72	33.8	29	1	US-08-152-488-10	Sequence 10, Appl
36	72	33.8	29	1	US-08-152-488-11	Sequence 11, Appl
37	72	33.8	29	1	US-08-303-025-10	Sequence 10, Appl
38	72	33.8	29	1	US-08-303-025-11	Sequence 11, Appl
39	72	33.8	29	1	US-08-303-025-13	Sequence 13, Appl
40	72	33.8	29	1	US-08-677-304-10	Sequence 10, Appl
41	72	33.8	29	1	US-08-677-304-11	Sequence 11, Appl
42	72	33.8	29	2	US-08-436-703B-3	Sequence 3, Appli
43	72	33.8	29	2	US-08-436-703B-15	Sequence 15, Appl
44	71.5	33.6	49	3	US-08-995-172-14	Sequence 14, Appl
45	71.5	33.6	49	3	US-08-839-624-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-405-743A-2
; Sequence 2, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A
; CURRENT APPLICATION NUMBER: US/09/405,743A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-405-743A-2

Query Match 100.0%; Score 213; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.2e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
```

RESULT 2
US-09-405-743A-7
; Sequence 7, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A
; CURRENT APPLICATION NUMBER: US/09/405,743A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-405-743A-7

```
  Query Match             64.8%;  Score 138;  DB 4;  Length 109;
  Best Local Similarity   76.6%;  Pred. No. 2.9e-08;
  Matches   36;  Conservative    0;  Mismatches    7;  Indels    4;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
              |||||| ||||  || | ||||||  |||  || | ||||||||||
Db         65 AKKYAKAAKAE--KKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109


RESULT 3
US-09-405-743A-3
; Sequence 3, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-3

  Query Match             63.1%;  Score 134.5;  DB 4;  Length 56;
  Best Local Similarity   68.4%;  Pred. No. 3.5e-08;
  Matches   39;  Conservative    0;  Mismatches    5;  Indels   13;  Gaps    4;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKK--AAKYEKAAAEKAAAKEAAYEA 45
              |||||||     |||||     ||||||||||   ||| || | ||||||||||
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAEAKY-KAEAAKAAAKEAAYEA 56


RESULT 4
US-09-405-743A-5
; Sequence 5, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 77
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-5

  Query Match             61.5%;  Score 131;  DB 4;  Length 77;
  Best Local Similarity   50.6%;  Pred. No. 1.1e-07;
  Matches   39;  Conservative    0;  Mismatches    6;  Indels   32;  Gaps    3;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||     |||||     |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 -KAAAEKAAAKEAAYEA 45
               || | ||||||||||||
Db         61 YKAEAAKAAAKEAAYEA 77


RESULT 5
US-09-405-743A-6
; Sequence 6, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 86
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-6

  Query Match             59.4%;  Score 126.5;  DB 4;  Length 86;
  Best Local Similarity   45.3%;  Pred. No. 3.8e-07;
  Matches   39;  Conservative    0;  Mismatches    6;  Indels   41;  Gaps    3;

Qy          1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE---------------------- 29
              |||||||     |||||     |||||||||| || |
Db          1 AKKYAKKEKAYAKKAEKAAKKAEAKAYKAAEAKKKAKAEAKKYAKAAKAEKKEYAAAEAK 60

Qy         30 ----------KAAAEKAAAKEAAYEA 45
                      || | ||||||||||||
Db         61 YKAEAAKKAYKAEAAKAAAKEAAYEA 86


RESULT 6
US-09-405-743A-4
; Sequence 4, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-4

  Query Match             56.6%;  Score 120.5;  DB 4;  Length 66;
  Best Local Similarity   71.1%;  Pred. No. 1.3e-06;
  Matches   32;  Conservative    0;  Mismatches    6;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |||||| ||||  || | |||||      || | ||||||||||||
Db         29 AKKYAKAAKAE--KKEYAAAEAK-----YKAEAAKAAAKEAAYEA 66


RESULT 7
US-09-405-743A-1
; Sequence 1, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
```

```
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 35
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:    SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-1

  Query Match             50.7%;  Score 108;  DB 4;  Length 35;
  Best Local Similarity   64.4%;  Pred. No. 1.4e-05;
  Matches   29;  Conservative    1;  Mismatches    5;  Indels   10;  Gaps    2;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||||||| ||  ||||||        |  || | :||||||||||
Db          1 AKKYAKKEKA--AKKAYK--------KEAKAKAAEAAAKEAAYEA 35


RESULT 8
US-09-095-855-201
; Sequence 201, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  Compounds and Methods for
;    TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/095,855
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/705,347
;      FILING DATE:  29-AUG-1996
;      APPLICATION NUMBER:  08/873,970
;      FILING DATE:  12-JUN-1997
;      APPLICATION NUMBER:  08/997,362
;      FILING DATE:  23-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  223 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-095-855-201

  Query Match             42.7%;  Score 91;  DB 3;  Length 223;
  Best Local Similarity   59.6%;  Pred. No. 0.0062;
  Matches   28;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKAAKYEKAA-AEKAAAKEAAYEA 45
              ||| | | | |:||  | ||| ||||   :||| |:|| ||:|| :|
Db        137 AKKAATKAAPAKKATAAKKAAPAKKATAAKKAAPAKKAPAKKAATKA 183


RESULT 9
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Tan, Paul L. J.
;  TITLE OF INVENTION: Compounds and Methods for Treatment and
;  TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
;  FILE REFERENCE: 11000.1002c4
;  CURRENT APPLICATION NUMBER: US/09/205,426
;  CURRENT FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: 09/095,855
;  EARLIER FILING DATE: 1998-06-11
;  EARLIER APPLICATION NUMBER: 08/997,362
;  EARLIER FILING DATE: 1997-12-23
;  EARLIER APPLICATION NUMBER: 08/873,970
;  EARLIER FILING DATE: 1997-06-12
;  EARLIER APPLICATION NUMBER: 08/705,347
;  EARLIER FILING DATE: 1996-08-29
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 201
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-09-205-426-201

  Query Match             42.7%;  Score 91;  DB 4;  Length 223;
  Best Local Similarity   59.6%;  Pred. No. 0.0062;
  Matches   28;  Conservative    5;  Mismatches   12;  Indels    2;  Gaps    2;

Qy          1 AKKYA-KKAKAEKAKKAYKAAEAKKAAKYEKAA-AEKAAAKEAAYEA 45
              ||| | | | |:||  | ||| ||||   :||| |:|| ||:|| :|
Db        137 AKKAATKAAPAKKATAAKKAAPAKKATAAKKAAPAKKAPAKKAATKA 183


RESULT 10
US-09-041-889-27
; Sequence 27, Application US/09041889
; Patent No. 6033864
;  GENERAL INFORMATION:
;    APPLICANT:  Braun, Jonathan
;    APPLICANT:  Cohavy, Offer
;    TITLE OF INVENTION:  Diagnosis, Prevention and Treatment of
;    TITLE OF INVENTION:  Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;    TITLE OF INVENTION:  Microbial UC pANCA antigens
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,889
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/837,058
;      FILING DATE:  11-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-PM 3006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  214 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-041-889-27

  Query Match             42.0%;  Score 89.5;  DB 3;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.0086;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  |  |    : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 11
US-09-417-264-27
; Sequence 27, Application US/09417264
; Patent No. 6537768
;  GENERAL INFORMATION:
;    APPLICANT:  Braun, Jonathan
;    APPLICANT:  Cohavy, Offer
;    TITLE OF INVENTION:  Diagnosis, Prevention and Treatment of
;    TITLE OF INVENTION:  Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;    TITLE OF INVENTION:  Microbial UC pANCA antigens
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/417,264
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/041,889
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-PM 3006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  214 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-417-264-27

  Query Match             42.0%;  Score 89.5;  DB 4;  Length 214;
  Best Local Similarity   55.6%;  Pred. No. 0.0086;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    3;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              ||| |||| |:|| || | |  |  |    : || || ||:|| :|
Db        111 AKKVAKKAPAKKATKAAKKAATKAPA---RKAATKAPAKKAATKA 152


RESULT 12
US-09-252-991A-29581
; Sequence 29581, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29581
;   LENGTH: 407
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29581

  Query Match             40.4%;  Score 86;  DB 4;  Length 407;
  Best Local Similarity   51.1%;  Pred. No. 0.039;
  Matches   24;  Conservative    6;  Mismatches   15;  Indels    2;  Gaps    1;

Qy          1 AKKYAKKAKAEKAKKA--YKAAEAKKAAKYEKAAAEKAAAKEAAYEA 45
              |:| |:  ||::||||   |||| || |   |  ||  | |:|| :|
Db        185 AQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKRAEDEAKKKAAEDA 231


RESULT 13
US-09-252-991A-32957
; Sequence 32957, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 32957
;   LENGTH: 316
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32957

  Query Match             36.6%;  Score 78;  DB 4;  Length 316;
  Best Local Similarity   54.8%;  Pred. No. 0.22;
  Matches   23;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          1 AKKYAKKAKAEKAKKAYKAAEAKKAAKYEKAAAEKAAAKEAA 42
              ||  || |    || | | | || |||    || | ||| ||
Db        157 AKAAAKPAAKPAAKPAAKTAAAKPAAKPAAKAAAKPAAKPAA 198


RESULT 14
US-09-107-532A-5094
```

```
; Sequence 5094, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 5094:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 700 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...700
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5094:
US-09-107-532A-5094

  Query Match             36.4%;  Score 77.5;  DB 4;  Length 700;
  Best Local Similarity   53.7%;  Pred. No. 0.55;
  Matches   22;  Conservative    7;  Mismatches    9;  Indels    3;  Gaps    2;

Qy          6 KKAKAEKAKKAYKAAEAKKAAKYE-KAAAEKAAAKEAAYEA 45
              |: | |: ||  || ||||||: : | ||:.||| :|| :|
Db        542 KEKKEEEEKK--KAEEAKKAAEEQAKQAAQAAAAAQAAQQA 580


RESULT 15
US-09-041-889-39
; Sequence 39, Application US/09041889
; Patent No. 6033864
;  GENERAL INFORMATION:
;    APPLICANT:  Braun, Jonathan
;    APPLICANT:  Cohavy, Offer
;    TITLE OF INVENTION:  Diagnosis, Prevention and Treatment of
;    TITLE OF INVENTION:  Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;    TITLE OF INVENTION:  Microbial UC pANCA antigens
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,889
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/837,058
;      FILING DATE:  11-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-PM 3006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  103 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-041-889-39

  Query Match             36.2%;  Score 77;  DB 3;  Length 103;
  Best Local Similarity   52.2%;  Pred. No. 0.088;
  Matches   24;  Conservative    1;  Mismatches   13;  Indels    8;  Gaps    1;

Qy          1 AKKYAKKAKAEKA--------KKAYKAAEAKKAAKYEKAAAEKAAA 38
              ||  |||| | ||         ||| ||| |||| |     |:| ||
Db         51 AKPKAKKAGAAKAKKPAGATPKKAKKAAGAKKAVKKTPKKAKKPAA 96


Search completed: January 21, 2004, 09:02:20
Job time : 8.47403 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:57:01 ; Search time 19.8182 Seconds
(without alignments)
528.928 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	189.5	36.5	421	2	JV0057	tolA protein - Esc
2	181.5	35.0	394	2	F90725	membrane spanning
3	181.5	35.0	394	2	G85576	membrane spanning
4	181	34.9	210	2	A25550	histone H1 - sea u
5	179	34.5	347	2	E83525	TolA protein PA097
6	178	34.3	376	2	AG0592	tolA protein [impo
7	178	34.3	388	2	AC0138	TolA colicin impor
8	175.5	33.8	344	2	S34153	mst101-1 protein -
9	172.5	33.2	211	2	A28100	histone H1-beta, e
10	171.5	33.0	1390	2	S51364	sperm tail-specifi
11	170.5	32.9	248	1	HSUR1P	histone H1, gonada
12	170.5	32.9	1701	2	T09127	probable erythrocy
13	165	31.8	220	2	A28456	histone H1.10 - ch
14	164.5	31.7	231	2	S59589	histone H1 - Chlam
15	164	31.6	265	2	S19113	cgcr-4 protein - C
16	164	31.6	311	2	T17698	hypothetical prote
17	163.5	31.5	384	2	B43592	outer membrane pro
18	163	31.4	182	2	S61926	histone H1 homolog
19	163	31.4	225	2	B28456	histone H1.11L - c
20	162.5	31.3	1128	2	T30296	R27-2 protein - Tr
21	162	31.2	356	2	A82152	tolA protein VC183
22	161.5	31.1	328	2	A44993	cytosolic repetiti
23	161	31.0	219	2	E60110	repetitive protein
24	160.5	30.9	206	1	HSTR1R	histone H1 - rainb
25	160.5	30.9	226	1	S51660	histone H1-5 [vali
26	159.5	30.7	219	2	C28456	histone H1.11R - c
27	158	30.4	219	1	HSHU1B	histone H1-4 [vali
28	157	30.3	771	1	A33430	h-caldesmon - chic
29	156.5	30.2	218	2	A23055	histone H1.01 - ch
30	154.5	29.8	217	2	S29309	hypothetical prote
31	154.5	29.8	241	2	JN0748	histone H1-II - Vo
32	154.5	29.8	309	2	G83013	polyhydroxyalkanoa
33	153.5	29.6	369	2	H71321	conserved hypothet
34	153.5	29.6	372	2	G64064	outer membrane int
35	152.5	29.4	224	2	D28456	histone H1.03 - ch
36	151	29.1	1403	2	T11583	probable translati
37	150	28.9	212	2	A28470	histone H1 - mouse
38	150	28.9	217	2	A26721	histone H1-gamma,
39	150	28.9	288	2	T06257	histone H1 (clone
40	148	28.5	218	2	S01262	histone H1 - musco
41	147.5	28.4	219	2	I49742	histone H1 - mouse
42	147	28.3	208	2	T23778	histone H1.1 - Cae
43	147	28.3	221	2	S49482	histone H1 - mouse
44	147	28.3	243	2	AE1689	hypothetical prote
45	147	28.3	340	2	A35630	regulatory protein

ALIGNMENTS

RESULT 1
JV0057
tolA protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002
C;Accession: JV0057; B64810
R;Levengood, S.K.; Webster, R.E.
J. Bacteriol. 171, 6600-6609, 1989
A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their produ
A;Reference number: JV0057; MUID:90078104; PMID:2687247
A;Accession: JV0057
A;Molecule type: DNA
A;Residues: 1-421 <LEV>
A;Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019
A;Experimental source: strain JM105
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64810
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-421 <BLAT>
A;Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73833.1; PID:g1786960;
A;Experimental source: strain K-12, substrain MG1655
C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach t
C;Genetics:
A;Gene: tolA
A;Map position: 17 min
A;Start codon: GTG
C;Keywords: nucleotide binding; P-loop; transmembrane protein
F;14-34/Domain: transmembrane #status predicted <MSS>
F;78-301/Domain: helical #status predicted <HSR>
F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 36.5%; Score 189.5; DB 2; Length 421;
Best Local Similarity 53.0%; Pred. No. 1.6e-05;
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5;

```
Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: :|||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADAKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||    |   |||  | |||||:|    |||:  |||| | || | |
Db        178 AAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAATEAAEKAKAEAEKKAAAEKA 232
```

RESULT 2
F90725

membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90725
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90725
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34197.1; PID:g13360233; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0774

Query Match 35.0%; Score 181.5; DB 2; Length 394;
Best Local Similarity 52.2%; Pred. No. 4.7e-05;
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

```
Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: : ||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADDKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||   |   |||  | |||||:|   || | | |  | ||||::||
Db        178 AAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAAAEKAAADKKAAEKAAAEKAA 232
```

RESULT 3
G85576
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C;Accession: G85576
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <STO>
A;Cross-references: GB:AE005174; NID:g12513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:Z09
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: tolA

Query Match 35.0%; Score 181.5; DB 2; Length 394;
Best Local Similarity 52.2%; Pred. No. 4.7e-05;
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

```
Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: : ||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADDKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||   |   |||  | |||||:|   || | | |  | ||||::||
Db        178 AAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAAAEKAAADKKAAEKAAAEKAA 232
```

RESULT 4
A25550
histone H1 - sea urchin (Lytechinus pictus)
C;Species: Lytechinus pictus (painted urchin)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
C;Accession: A25550
R;Knowles, J.A.; Childs, G.J.
Nucleic Acids Res. 14, 8121-8133, 1986
A;Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus an
A;Reference number: A25550; MUID:87040778; PMID:3022245
A;Accession: A25550
A;Molecule type: DNA
A;Residues: 1-210 <KNO>
A;Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617
C;Superfamily: histone H1
C;Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 34.9%; Score 181; DB 2; Length 210;
Best Local Similarity 50.0%; Pred. No. 3.2e-05;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

```
Qy          2 KKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAE----AKKKAKAEAKKYAKEAAKAK 57
              |  |:|| :| ||||| | :||   :|:|   || :    ||| ||  |||  | |||||
Db         96 KTEAQKA-RAAAKKAKLAAKKKEQKEKKAAKTKARKEKLAAKKAAKKAAKKVKKPAAKAK 154

Qy         58 KEAYKAEAKKYAKAAKAEKKEYAAAEAKKA-------EAAKAYKAEAAKAAAKEAA 106
              | | ||  |  ||  || ||  |   ||||       :|||  |  | | |||:||
Db        155 KPAKKAAKKPAAK--KAAKKPAAKKPAKKAAKKPAAKKAAKPAKKAAKKPAAKKAA 208
```

RESULT 5
E83525
TolA protein PA0971 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83525
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-347 <STO>
A;Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: tolA; PA0971

Query Match 34.5%; Score 179; DB 2; Length 347;
Best Local Similarity 49.1%; Pred. No. 6.2e-05;
Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;

```
Qy          6 KKAEKAYAKKAKAAKEK------KAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              :| |:     ||||::|      || |:| |:| || |||| |:|:| :  |:|  |||
Db         99 QKLEQQQVAAAKAAEQKKADEARKAEAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKR 158

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              | : ||||  |||:  ||: || :|||  | :| |  ||:|| |:|| ||
Db        159 A-EDEAKK--KAAEDAKKK-AAEDAKKKAAEEAKKKAAAEAAKKKAAVEA 204
```

RESULT 6
AG0592
tolA protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AG0592
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AG0592
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <PAR>

```
A;Cross-references: GB:AL513382; PIDN:CAD05209.1; PID:g16501979; GSPDB:GN00176
C;Genetics:
A;Gene: STY0793

  Query Match             34.3%;  Score 178;  DB 2;  Length 376;
  Best Local Similarity   52.3%;  Pred. No. 7.6e-05;
  Matches   58;  Conservative   10;  Mismatches   33;  Indels   10;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||  |   :|| |: |||| : |  |: || |  ||:|||||:|||   || ||:|||
Db        137 AKAAADAKKKAEAEAAKAAADAKKKAEAEA-AKAAADAKKKAEAEA---AKAAAEAKK-- 190

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAA---KEAAYE 108
               ||||:    || |:||  | |    ||| |   | ||||||   |:|| |
Db        191 -KAEAEAAKAAADAKKKADAEAAKAAAEAKKKADAAAAKAAADAKKKAAAE 240


RESULT 7
AC0138
TolA colicin import membrane protein [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0138
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0138
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-388 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89966.1; PID:g15979190; GSPDB:GN00175
C;Genetics:
A;Gene: tolA

  Query Match             34.3%;  Score 178;  DB 2;  Length 388;
  Best Local Similarity   52.1%;  Pred. No. 7.8e-05;
  Matches   63;  Conservative   12;  Mismatches   28;  Indels   18;  Gaps    7;

Qy          2 KKYAKKAEKAYAKKAKAAKEKK------AYAKKEA-KAYKA-AEAKKKAKAEAKKYAKEA 53
              |: |:: : | |  ||| :|:|      | || || |  || |||:|||:|||| |  |
Db        140 KQAAEQQKIAAAAVAKAKEEQKQAETAAAQAKAEADKIVKAQAEAQKKAEAEAKKEAAVA 199

Qy         54 AKAKKEAYKAEAKKYAKAAK------AEKKEYAAAEAKKAEAAK--AYKAEAAKAAAKEA 105
              | |||:|  |:|||  : |:      ||||  | || ||| |||  |  ||| | || ||
Db        200 AAAKKQA-DADAKKAVEVAEKAAADAAEKKAAADAE-KKAAAAKKVAAAAEAKKKAAAEA 257

Qy        106 A 106
              |
Db        258 A 258


RESULT 8
S34153
mst101-1 protein - fruit fly (Drosophila hydei)
C;Species: Drosophila hydei
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C;Accession: S34153
R;Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A;Reference number: S34153
A;Accession: S34153
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-344 <NEE>
A;Cross-references: EMBL:X73480; NID:g313199; PID:g313200
C;Genetics:
A;Gene: FlyBase:Dhyd/mst101
A;Cross-references: FlyBase:FBgn0011816
C;Superfamily: neurofilament triplet H protein

  Query Match             33.8%;  Score 175.5;  DB 2;  Length 344;
  Best Local Similarity   49.2%;  Pred. No. 0.0001;
  Matches   58;  Conservative    8;  Mismatches   35;  Indels   17;  Gaps    5;

Qy          1 AKKYAKKAEK----AYAKKAKAAKEKKAYAKKEAKAYKAAE-------AKKKAKAEAKKY 49
              |||  : |||       ||| | | |||  |:    |  :|||       |||: :|  ||
Db         65 AKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKK 124

Qy         50 AKEAAKAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKK-AEAAKAYKAEAAKAAAKEAA 106
                |||| :|||  || || |:||| ||:   ||| || |||||  |  | |    |||
Db        125 CAEAAKKEKEA--AEKKKCAEAAKKEKE---AAEKKKCAEAAKKEKEAAEKKKCAEAA 177


RESULT 9
A28100
histone H1-beta, embryonic - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 23-Feb-1997
C;Accession: A28100
R;Lai, Z.C.; Childs, G.
Mol. Cell. Biol. 8, 1842-1844, 1988
A;Title: Characterization of the structure and transcriptional patterns of the gene enco
A;Reference number: A28100; MUID:88246461; PMID:2837660
A;Accession: A28100
A;Molecule type: DNA
A;Residues: 1-211 <LAI>
A;Cross-references: GB:M20314
C;Superfamily: histone H1
C;Keywords: chromosomal protein; DNA binding; embryo; nucleosome; nucleus

  Query Match             33.2%;  Score 172.5;  DB 2;  Length 211;
  Best Local Similarity   48.7%;  Pred. No. 0.00011;
  Matches   55;  Conservative    9;  Mismatches   40;  Indels    9;  Gaps    4;

Qy          3 KYAKKAE-KAYAKKAK--AAKEKKAYAKKEAKAYKAAEAK-KKAKAEAKKYAKEAAK--- 55
              |  || | |: |:||:  | | | |  |||||  ||| :| || |  ||| :|:  |
Db         88 KLGKKKEGKSDAQKARDAAKKAKLAAKKKEAKEKKAARSKAKKEKLAAKKASKKTTKKVK 147

Qy         56 --AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
                | |:| |  ||| ||   |:|     | || | | || |  ||| |||: |
Db        148 KPAAKKAKKPAAKKAAKKPAAKKPAAKKAAAKPAPAKKAAKKPAAKKAAKKVA 200


RESULT 10
S51364
sperm tail-specific protein mst101(2) - fruit fly (Drosophila hydei)
C;Species: Drosophila hydei
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S51364; S34154
R;Neesen, J.; Padmanabhan, S.; Buenemann, H.
Eur. J. Biochem. 225, 1089-1095, 1994
A;Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represe
lpha-helical rods within the extremely elongated spermatozoa of Drosophila hydei.
A;Reference number: S51364; MUID:95045538; PMID:7957199
A;Accession: S51364
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1390 <NEE>
A;Cross-references: EMBL:X73481
R;Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A;Reference number: S34153
A;Accession: S34154
A;Molecule type: DNA
A;Residues: 1-163,'E',164-236,'Q',237-254,257-320,'E',321-1390 <NEW>
A;Cross-references: EMBL:X73481; NID:g313201; PID:g313202
C;Genetics:
A;Gene: mst101(2)
A;Cross-references: FlyBase:FBgn0011816

  Query Match             33.0%;  Score 171.5;  DB 2;  Length 1390;
  Best Local Similarity   46.2%;  Pred. No. 0.00051;
  Matches   54;  Conservative   11;  Mismatches   37;  Indels   15;  Gaps    4;
```

```
Qy          1 AKKYAKKAEKAY----AKKAKAAKEKKAYAKKEAKAYKAAE-------AKKKAKAEAKKY 49
              |||  : |||       ||| | | |||   ::  |  :|||       |||: :|  ||
Db       1067 AKKLKEAAEKKQCEERAKKEKEAAEKKQCEERAKKLKEAAEKKQCEERAKKEKEAAEKKR 1126

Qy         50 AKEAAKAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
               :|||| :|||  || || |:||| ||:  |  : | |||||  |  | |    |||
Db       1127 CEEAAKREKEA--AEKKKCAEAAKKEKE--ATEKQKCAEAAKKEKEAAEKKKCAEAA 1179


RESULT 11
HSUR1P
histone H1, gonadal - sea urchin (Parechinus angulosus)
C;Species: Parechinus angulosus (angulate urchin)
C;Date: 31-Mar-1980 #sequence_revision 31-Mar-1980 #text_change 16-Feb-1997
C;Accession: A91090; A91091; A02586
R;Strickland, W.N.; Strickland, M.; de Groot, P.C.; von Holt, C.; Wittmann-Liebold, B.
Eur. J. Biochem. 104, 559-566, 1980
A;Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus ang
rminal cyanogen bromide peptides.
A;Reference number: A91090; MUID:80156831; PMID:6767609
A;Contents: sequence of residues 1-84
A;Accession: A91090
A;Molecule type: protein
A;Residues: 1-248 <STR>
R;Strickland, W.N.; Strickland, M.; Brandt, W.F.; von Holt, C.; Lehmann, A.; Wittmann-Li
Eur. J. Biochem. 104, 567-578, 1980
A;Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus ang
A;Reference number: A91091; MUID:80156832; PMID:7363905
A;Accession: A91091
A;Molecule type: protein
A;Residues: 80-248 <ST2>
A;Note: 144-Arg was also found
C;Superfamily: histone H1
C;Keywords: DNA binding; nucleosome; sperm

  Query Match           32.9%;  Score 170.5;  DB 1;  Length 248;
  Best Local Similarity   46.0%;  Pred. No. 0.00017;
  Matches   52;  Conservative    9;  Mismatches   41;  Indels   11;  Gaps    2;

Qy          2 KKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAY 61
              ||  | :  | |||||||  |||   | |   ||| |||||  |  :| | :| ||||
Db        120 KKAKKTSAAAKAKKAKAAAAKKARKAKAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKK 179

Qy         62 KA--EAKKYAK---------AAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  :||| ||         | |: ||: |    |||: |   :  ||| | :
Db        180 KAAKKAKKPAKKSPKKAKKPAKKSPKKKKAKRSPKKAKKAAGKRKPAAKKARR 232


RESULT 12
T09127
probable erythrocyte-binding protein MAEBL - Plasmodium yoelii
C;Species: Plasmodium yoelii
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T09127
R;Kappe, S.H.I.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998
A;Title: A family of chimeric erythrocyte binding proteins of malaria parasites.
A;Reference number: Z16577; MUID:98115903; PMID:9448314
A;Accession: T09127
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1701 <KAP>
A;Cross-references: EMBL:AF031886; NID:g2947227; PID:g2947228
A;Experimental source: subspecies yoelii; strain YM
C;Genetics:
A;Gene: maebl
A;Introns: 62/1; 1648/1; 1674/2; 1697/1
C;Keywords: alternative splicing; cell binding; erythrocyte invasion

  Query Match           32.9%;  Score 170.5;  DB 2;  Length 1701;
  Best Local Similarity   53.5%;  Pred. No. 0.00068;
  Matches   54;  Conservative   11;  Mismatches   31;  Indels    5;  Gaps    5;

Qy          5 AKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKA- 63
              |||||:   |||:| | |   |||:|:| | || :|| |||| | | |  | : | ||
Db       1205 AKKAEEE-RKKAEAVK-KAEEAKKKAEAAKKAEERKK-KAEAAKKALERKKKSEAAKKAL 1261

Qy         64 EAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              | || |:|||  :::  |  |||||  |  |||||| | :|
Db       1262 ERKKKAEAAKKAEEKKKAEAAKKAEEEKK-KAEAAKKAEEE 1301


RESULT 13
A28456
histone H1.10 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 04-Mar-2000
C;Accession: A28456
R;Coles, L.S.; Robins, A.J.; Madley, L.K.; Wells, J.R.E.
J. Biol. Chem. 262, 9656-9663, 1987
A;Title: Characterization of the chicken histone H1 gene complement. Generation of a com
A;Reference number: A92655; MUID:87250632; PMID:3597432
A;Accession: A28456
A;Molecule type: DNA
A;Residues: 1-220 <COL>
A;Cross-references: GB:M17018; NID:g211834; PIDN:AAA48788.1; PID:g211835
C;Superfamily: histone H1
C;Keywords: acetylated amino end; chromosomal protein; DNA binding; nucleosome; nucleus
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

  Query Match           31.8%;  Score 165;  DB 2;  Length 220;
  Best Local Similarity   46.4%;  Pred. No. 0.00034;
  Matches   52;  Conservative   11;  Mismatches   41;  Indels    8;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEA----KAYKAAEAK---KKAKAEAKKYAKEA 53
              :||  :  |||  |:  ||| ||  ||| |    |  ||| ||   |||| |    |:|
Db        109 SKKPGEVKEKAPRKRTPAAKPKKPAAKKPASAAKKPKKAAAAKKSPKKAKKPAAAATKKA 168

Qy         54 AKAKKEAYKAEAKKYAKAAKAEKKEYAA-AEAKKAEAAKAYKAEAAKAAAKE 104
              ||: |:| ||   | |  ||:  |  |   :| | :|||   |:| |||||:
Db        169 AKSPKKATKAAKPKKAATAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAAKK 220


RESULT 14
S59589
histone H1 - Chlamydomonas reinhardtii
C;Species: Chlamydomonas reinhardtii
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 23-Jul-1999
C;Accession: S59589; S62122
R;Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.
Curr. Genet. 28, 333-345, 1995
A;Title: The organization structure and regulatory elements of Chlamydomonas histone gen
A;Reference number: S59581; MUID:96120862; PMID:8590479
A;Accession: S59589
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-231 <FAB>
A;Cross-references: EMBL:U16726
A;Note: the authors did not translate the codon for residue 1
R;Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.
submitted to the EMBL Data Library, October 1994
A;Description: The organization, structure and controlling elements of Chlamydomonas his
A;Reference number: S62122
A;Accession: S62122
A;Molecule type: DNA
A;Residues: 1-173,'P',174-231 <FAW>
A;Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98452.1; PID:g571480
C;Genetics:
A;Introns: 62/3; 101/3
C;Superfamily: histone H1
C;Keywords: chromosomal protein; DNA binding; nucleosome

  Query Match           31.7%;  Score 164.5;  DB 2;  Length 231;
  Best Local Similarity   45.9%;  Pred. No. 0.00038;
  Matches   56;  Conservative   10;  Mismatches   39;  Indels   17;  Gaps    5;
```

```
Qy          2 KKYAKKA---EKAYAKKAKAAKEKKAYAKKEA----KAYKAAEAKKKAKAEAKKYAKEAA 54
              || ||||   :|| | | : | :|    |||     |: | || | | : | || || |
Db        100 KKAAKKAATPKKAAAPKKEGAVKKTKAPKKEGEKKPKSAKKAEKKPKKEGEKKKAAKPAK 159

Qy         55 KAKKEAYKAEAKKYAKAA----KAEKKEYAAAE---AKKAEAAK---AYKAEAAKAAAKE 104
                ||      |||  | |    |||||  |||:    || :|||     ||:||| |||:
Db        160 AEKKPKAAKPAKKTTKKAAAKPKAEKKPKAAAKPKAEKKPKAAKPKAEKKAKAAKPAAKK 219

Qy        105 AA 106
              :|
Db        220 SA 221


RESULT 15
S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C;Accession: S19113; S14466
R;Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A;Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A;Reference number: S19113; MUID:92119224; PMID:1731966
A;Accession: S19113
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <WAK>
A;Cross-references: EMBL:X17208; NID:g18136; PIDN:CAA35080.1; PID:g18137
C;Genetics:
A;Gene: cgcr-4

  Query Match             31.6%;  Score 164;  DB 2;  Length 265;
  Best Local Similarity   39.4%;  Pred. No. 0.00045;
  Matches   50;  Conservative   21;  Mismatches   38;  Indels   18;  Gaps    2;

Qy          1 AKKYAKKAEKA--------------YAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEA 46
              |:  ||:|::|                | :|:|  | :|:|  ||:|: ||| : :| |:|
Db         79 AEAKAKEADEARAAEAEARAVAEAWAAAEAEARAEAEAWAAAEAEAWAAAEVEARAAAQA 138

Qy         47 KKYAKEAAKAKKEAYKAEAKKYAKAAK----AEKKEYAAAEAKKAEAAKAYKAEAAKAAA 102
              :  |:  | |  ||  | | : | |||    || :  |||||: | ||:|  |  |:|||
Db        139 RVAAEARAAAAAEARAAAAAEAAAAAKARVAAEARAAAAAEARAAAAAEAAAAAKARAAA 198

Qy        103 KEAAYEA 109
              :  | ||
Db        199 EAKAKEA 205


Search completed: January 21, 2004, 09:01:40
Job time : 20.8182 secs
```

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:56:26 ; Search time 14.8636 Seconds
(without alignments)
344.863 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	189.5	36.5	421	1	TOLA_ECOLI	P19934 escherichia
2	181	34.9	210	1	H1_LYTPI	P06144 lytechinus
3	179	34.5	347	1	TOLA_PSEAE	P50600 pseudomonas
4	175.5	33.8	344	1	MST1_DROHY	Q08695 drosophila
5	171.5	33.0	1391	1	MST2_DROHY	Q08696 drosophila
6	170.5	32.9	248	1	H1_PARAN	P02256 parechinus
7	169.5	32.7	211	1	H1B_STRPU	P15869 strongyloce
8	166.5	32.1	668	1	FAU_DROME	Q9vgx3 drosophila
9	165	31.8	219	1	H110_CHICK	P08286 gallus gall
10	163.5	31.5	384	1	TMPB_TREPH	P29720 treponema p
11	163	31.4	224	1	H11L_CHICK	P08287 gallus gall
12	160.5	30.9	206	1	H1_ONCMY	P06350 oncorhynchu
13	160.5	30.9	225	1	H15_HUMAN	P16401 homo sapien
14	160	30.8	208	1	DBH_MYCSM	Q9zhc5 mycobacteri
15	159.5	30.7	218	1	H11R_CHICK	P08288 gallus gall
16	158	30.4	218	1	H14_HUMAN	P10412 homo sapien
17	157	30.3	771	1	CALD_CHICK	P12957 gallus gall
18	156.5	30.2	218	1	H101_CHICK	P08284 gallus gall
19	154.5	29.8	240	1	H12_VOLCA	Q08865 volvox cart
20	153.5	29.6	372	1	TOLA_HAEIN	P44678 haemophilus
21	152.5	29.4	223	1	H103_CHICK	P08285 gallus gall
22	151	29.1	1403	1	YDF3_SCHPO	Q10475 schizosacch
23	150	28.9	211	1	H12_MOUSE	P15864 mus musculu
24	150	28.9	217	1	H1G_STRPU	P07796 strongyloce
25	148	28.5	139	1	ASR_KLEPN	Q93mh5 klebsiella
26	148	28.5	217	1	H1_ANAPL	P09426 anas platyr
27	147.5	28.4	218	1	H14_MOUSE	P43274 mus musculu
28	147	28.3	207	1	H11_CAEEL	P10771 caenorhabdi
29	147	28.3	220	1	H13_MOUSE	P43277 mus musculu
30	146	28.1	217	1	H1_CHICK	P09987 gallus gall
31	146	28.1	299	1	RL22_DROME	P50887 drosophila
32	144	27.7	232	1	H12_GLYBA	P40264 glyptotendi
33	144	27.7	352	1	ALGP_PSEAE	P15276 pseudomonas
34	143.5	27.6	218	1	H12_RAT	P15865 rattus norv
35	143	27.6	214	1	DBH_MYCTU	P95109 mycobacteri
36	143	27.6	236	1	H1_NEUCR	Q8j0u2 neurospora
37	142.5	27.5	221	1	H1C_CHITE	P40277 chironomus
38	142	27.4	194	1	H1_SALTR	P02254 salmo trutt
39	141.5	27.3	232	1	H1B_CHITE	P40276 chironomus
40	141.5	27.3	346	1	RS6_AEDAE	Q9u761 aedes aegyp
41	140	27.0	232	1	H1A_CHITE	P40275 chironomus
42	139	26.8	213	1	H13_RABIT	P02251 oryctolagus
43	139	26.8	226	1	BASP_HUMAN	P80723 homo sapien
44	139	26.8	260	1	H11_VOLCA	Q08864 volvox cart
45	138.5	26.7	1020	1	NFH_HUMAN	P12036 homo sapien

ALIGNMENTS

RESULT 1
TOLA_ECOLI
ID TOLA_ECOLI STANDARD; PRT; 421 AA.
AC P19934;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TolA protein.
GN TOLA OR CIM OR EXCC OR LKY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JM105;
RX MEDLINE=90078104; PubMed=2687247;
RA Levengood S.K., Webster R.E.;
RT "Nucleotide sequences of the tolA and tolB genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP DOMAINS.
RX MEDLINE=91296736; PubMed=2068069;
RA Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT "TolA: a membrane protein involved in colicin uptake contains an
RT extended helical region.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN [5]
RP INTERACTION WITH PORINS.
RX MEDLINE=97133271; PubMed=8978668;
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

```
RA   Lloubes R.;
RT   "TolA central domain interacts with Escherichia coli porins.";
RL   EMBO J. 15:6408-6415(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX   MEDLINE=99332679; PubMed=10404600;
RA   Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT   "Filamentous phage infection: crystal structure of g3p in complex
RT   with its coreceptor, the C-terminal domain of TolA.";
RL   Structure 7:711-722(1999).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC       COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC       COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC       BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC       OF BACTERIOPHAGE DNA.
CC   -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC       AND LAMB.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28232; AAA24683.1; -.
DR   EMBL; AE000177; AAC73833.1; -.
DR   EMBL; D90713; BAA35405.1; -.
DR   PIR; JV0057; JV0057.
DR   PDB; 1TOL; 20-MAY-99.
DR   EcoGene; EG11007; tolA.
KW   Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW   Repeat; Inner membrane; 3D-structure; Complete proteome.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35    421       PERIPLASMIC (POTENTIAL).
FT   DOMAIN       48    310*      DOMAIN II (ALPHA-HELICAL).
FT   DOMAIN      311    421       DOMAIN III (FUNCTIONAL).
FT   DOMAIN      224    278       10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT                                A(2,4).
FT   HELIX       335    349
FT   TURN        350    351
FT   TURN        353    354
FT   HELIX       355    358
FT   TURN        359    360
FT   STRAND      363    369
FT   TURN        371    372
FT   STRAND      375    383
FT   HELIX       385    397
FT   HELIX       406    412
FT   TURN        413    414
FT   STRAND      416    421
SQ   SEQUENCE   421 AA;  43156 MW;  8B2F52B4B97C655E CRC64;

  Query Match             36.5%;  Score 189.5;  DB 1;  Length 421;
  Best Local Similarity   53.0%;  Pred. No. 4.6e-06;
  Matches   61;  Conservative   11;  Mismatches   32;  Indels   11;  Gaps    5;

Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: :|||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADAKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||   |   |||  | |||||:|    |||:  |||| | || | |
Db        178 AAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAATEAAEKAKAEAEKKAAAEKA 232


RESULT 2
H1_LYTPI
ID   H1_LYTPI       STANDARD;      PRT;   210 AA.
AC   P06144;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Late histone H1.
OS   Lytechinus pictus (Painted sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae;
OC   Lytechinus.
OX   NCBI_TaxID=7653;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=87040778; PubMed=3022245;
RA   Knowles J.A., Childs G.J.;
RT   "Comparison of the late H1 histone genes of the sea urchins
RT   Lytechinus pictus and Strongelocentrotus purpuratus.";
RL   Nucleic Acids Res. 14:8121-8133(1986).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04488; CAA28177.1; -.
DR   PIR; A25550; A25550.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   210 AA;  21746 MW;  08C38F64894007E2 CRC64;

  Query Match             34.9%;  Score 181;  DB 1;  Length 210;
  Best Local Similarity   50.0%;  Pred. No. 9.3e-06;
  Matches   58;  Conservative    8;  Mismatches   36;  Indels   14;  Gaps    4;

Qy          2 KKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAE----AKKKAKAEAKKYAKEAAKAK 57
              |  |:|| :| ||||| | :||   :|:|   || :    ||| ||  |||  | |||||
Db         96 KTEAQKA-RAAAKKAKLAAKKKEQKEKKAAKTKARKEKLAAKKAAKKAAKKVKKPAAKAK 154

Qy         58 KEAYKAEAKKYAKAAKAEKKEYAAAEAKKA-------EAAKAYKAEAAKAAAKEAA 106
              | | ||  |  ||  || ||  |   ||||       :|||  |  | | |||:||
Db        155 KPAKKAAKKPAAK--KAAKKPAAKKPAKKAAKKPAAKKAAKPAKKAAKKPAAKKAA 208


RESULT 3
TOLA_PSEAE
ID   TOLA_PSEAE     STANDARD;      PRT;   347 AA.
AC   P50600;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TolA protein.
GN   TOLA OR PA0971.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO;
RX   MEDLINE=97113525; PubMed=8955385;
```

```
RA   Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT   "Identification and characterization of the tolQRA genes of
RT   Pseudomonas aeruginosa.";
RL   J. Bacteriol. 178:7059-7068(1996).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RA   Duan K., Sokol P.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC       (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39558; AAC44660.2; -.
DR   EMBL; AE004530; AAG04360.1; -.
DR   PIR; E83525; E83525.
DR   InterPro; IPR006260; TonB_C.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
KW   Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW   Complete proteome.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     37       POTENTIAL.
FT   DOMAIN       38    347       PERIPLASMIC (POTENTIAL).
FT   DOMAIN      209    216       POLY-ALA.
SQ   SEQUENCE   347 AA;  37935 MW;  EEDD4B04AA095945 CRC64;

  Query Match             34.5%;  Score 179;  DB 1;  Length 347;
  Best Local Similarity   49.1%;  Pred. No. 1.9e-05;
  Matches   54;  Conservative   17;  Mismatches   29;  Indels   10;  Gaps    4;

Qy          6 KKAEKAYAKKAKAAKEK------KAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              :| |:     ||||::|      || |:| |:| || |||| |:|:| :  |:|  |||
Db         99 QKLEQQQVAAAKAAEQKKADEARKAEAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKR 158

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
                | : ||||  |||:  ||: || :|||  | :| |  ||:|| |:|| ||
Db        159 A-EDEAKK--KAAEDAKKK-AAEDAKKKAAAEEAKKKAAAEAAKKKAAVEA 204


RESULT 4
MST1_DROHY
ID   MST1_DROHY     STANDARD;      PRT;   344 AA.
AC   Q08695;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Axoneme-associated protein mst101(1).
GN   MST101(1).
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=94200512; PubMed=8150205;
RA   Neesen J., Buenemann H., Heinlein U.A.;
RT   "The Drosophila hydei gene Dhmst101(1) encodes a testis-specific,
RT   repetitive, axoneme-associated protein with differential abundance in
RT   Y chromosomal deletion mutant flies.";
RL   Dev. Biol. 162:414-425(1994).
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC       ASSOCIATED WITH AXONEMAL STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC       SPERMATID BUNDLES.
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73480; CAA51875.1; -.
DR   PIR; S34153; S34153.
DR   FlyBase; FBgn0011816; Dhyd\mst101(1).
KW   Sperm; Repeat; Multigene family.
FT   DOMAIN       58    337       19 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.
SQ   SEQUENCE   344 AA;  37793 MW;  24C65D2510387E2A CRC64;

  Query Match             33.8%;  Score 175.5;  DB 1;  Length 344;
  Best Local Similarity   49.2%;  Pred. No. 3.1e-05;
  Matches   58;  Conservative    8;  Mismatches   35;  Indels   17;  Gaps    5;

Qy          1 AKKYAKKAEK----AYAKKAKAAKEKKAYAKKEAKAYKAAE-------AKKKAKAEAKKY 49
              |||  : |||      |||| | | |||  |:   |  :|||       |||: :|  ||
Db         65 AKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKKCAEAAKKEKEAAEKKK 124

Qy         50 AKEAAKAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKK-AEAAKAYKAEAAKAAAKEAA 106
                |||| :|||  || || |:||| ||:   ||| || ||||| |  | |    |||
Db        125 CAEAAKKEKEA--AEKKKCAEAAKKEKE---AAEKKKCAEAAKKEKEAAEKKKCAEAA 177


RESULT 5
MST2_DROHY
ID   MST2_DROHY     STANDARD;      PRT;  1391 AA.
AC   Q08696;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Axoneme-associated protein mst101(2).
GN   MST101(2).
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=95045538; PubMed=7957199;
RA   Neesen J., Padmanabhan S., Buenemann H.;
RT   "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT   motif representing the major component of the sperm-tail-specific
RT   axoneme-associated protein family Dhmst101 form extended
RT   alpha-helical rods within the extremely elongated spermatozoa of
RT   Drosophila hydei.";
RL   Eur. J. Biochem. 225:1089-1095(1994).
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
```

```
CC   -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC       SPERMATIDS.
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC   -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC       STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC       REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73481; CAA51876.1; -.
DR   PIR; S51364; S51364.
DR   FlyBase; FBgn0020733; Dhyd\mst101(2).
KW   Sperm; Repeat; Multigene family; Polymorphism.
FT   DOMAIN      332   1268       59 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                [KR]-K-X-C-X-X-X-A-K-X-X-K-X-X-E.
SQ   SEQUENCE   1391 AA;  159000 MW;  1B2A368F30F48878 CRC64;

  Query Match             33.0%;  Score 171.5;  DB 1;  Length 1391;
  Best Local Similarity   46.2%;  Pred. No. 0.00017;
  Matches   54;  Conservative   11;  Mismatches   37;  Indels   15;  Gaps    4;

Qy          1 AKKYAKKAEKAY----AKKAKAAKEKKAYAKKEAKAYKAAE-------AKKKAKAEAKKY 49
              |||  : |||      ||| | | |||   ::  |  :|||        |||: :|  ||
Db       1068 AKKLKEAAEKKQCEERAKKEKEAAEKKQCEERAKKLKEAAEKKQCEERAKKEKEAAEKKR 1127

Qy         50 AKEAAKAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
                :|||| :|||  || || |:||| ||:  |  : | |||||  |  | |    |||
Db       1128 CEEAAKREKEA--AEKKKCAEAAKKEKE--ATEKQKCAEAAKKEKEAAEKKKCAEAA 1180


RESULT 6
H1_PARAN
ID   H1_PARAN       STANDARD;      PRT;   248 AA.
AC   P02256;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1, gonadal.
OS   Parechinus angulosus (Angulate sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
OC   Parechinus.
OX   NCBI_TaxID=7658;
RN   [1]
RP   SEQUENCE OF 1-84.
RX   MEDLINE=80156831; PubMed=6767609;
RA   Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT   protein and the sequence of amino acids in the four N-terminal
RT   cyanogen bromide peptides.";
RL   Eur. J. Biochem. 104:559-566(1980).
RN   [2]
RP   SEQUENCE OF 80-248.
RX   MEDLINE=80156832; PubMed=7363905;
RA   Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT   the entire primary structure.";
RL   Eur. J. Biochem. 104:567-578(1980).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: SPERM.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR   PIR; A91090; HSUR1P.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Sperm.
FT   VARIANT     144    144       K -> R.
SQ   SEQUENCE   248 AA;  26387 MW;  1B25B3F136541947 CRC64;

  Query Match             32.9%;  Score 170.5;  DB 1;  Length 248;
  Best Local Similarity   46.0%;  Pred. No. 5e-05;
  Matches   52;  Conservative    9;  Mismatches   41;  Indels   11;  Gaps    2;

Qy          2 KKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAY 61
              ||  | :  | |||||||  |||   | |   ||| |||||  |  :| | :| ||||
Db        120 KKAKKTSAAAKAKKAKAAAAKKARKAKAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKK 179

Qy         62 KA--EAKKYAK---------AAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  :||| ||          | |: ||: |    |||: |   :  ||| | :
Db        180 KAAKKAKKPAKKSPKKAKKPAKKSPKKKKAKRSPKKAKKAAGKRKPAAKKARR 232


RESULT 7
H1B_STRPU
ID   H1B_STRPU      STANDARD;      PRT;   211 AA.
AC   P15869;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1-beta, late embryonic.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88246461; PubMed=2837660;
RA   Lai Z.-C., Childs G.;
RT   "Characterization of the structure and transcriptional patterns of
RT   the gene encoding the late histone subtype H1-beta of the sea urchin
RT   Strongylocentrotus purpuratus.";
RL   Mol. Cell. Biol. 8:1842-1844(1988).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20314; AAA30052.1; -.
DR   PIR; A28100; A28100.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   211 AA;  22169 MW;  9F214581334BBE7A CRC64;
```

```
  Query Match             32.7%;  Score 169.5;  DB 1;  Length 211;
  Best Local Similarity   48.7%;  Pred. No. 5.1e-05;
  Matches   55;  Conservative    8;  Mismatches   41;  Indels    9;  Gaps    4;

Qy          3 KYAKKAE-KAYAKKA--KAAKEKKAYAKKEAKAYKAAEAK-KKAKAEAKKYAKEAAK--- 55
              |  || | |: |:||   | | | |  |||||  ||| :| || |  ||| :|:  |
Db         88 KLGKKKEGKSDAQKAPDAAKKAKLAAKKKEAKEKKAARSKAKKEKLAAKKASKKTTKKVK 147

Qy         56 --AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
                | |:| |  ||| ||   |:|     | || | | || |  ||| |||: |
Db        148 KPAAKKAKKPAAKKAAKKPAAKKPAAKKAAAKPAPAKKAAKKPAAKKAAKKVA 200


RESULT 8
FAU_DROME
ID   FAU_DROME      STANDARD;      PRT;   668 AA.
AC   Q9VGX3; Q95S18; Q9VGX1; Q9VGX2; Q9Y0F9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Anoxia upregulated protein.
GN   FAU OR CG6544.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S; TISSUE=Head;
RX   MEDLINE=99097004; PubMed=9878744;
RA   Ma E., Xu T., Haddad G.G.;
RT   "Gene regulation by O2 deprivation: an anoxia-regulated novel gene in
RT   Drosophila melanogaster.";
RL   Brain Res. Mol. Brain Res. 63:217-224(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM E).
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: Plays an important role in the regulation of tissue
CC       responsiveness to oxygen deprivation.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=A;
CC         IsoId=Q9VGX3-1; Sequence=Displayed;
CC       Name=B;
CC         IsoId=Q9VGX3-2; Sequence=VSP_004048, VSP_004049;
CC       Name=C;
CC         IsoId=Q9VGX3-3; Sequence=VSP_004046, VSP_004047;
CC       Name=D;
CC         IsoId=Q9VGX3-4; Sequence=VSP_004050, VSP_004051;
CC       Name=E;
CC         IsoId=Q9VGX3-5; Sequence=VSP_004052;
CC   -!- TISSUE SPECIFICITY: Concentrated in lamina neurons, first optic
CC       lobe neurons and cortical neurons of central brain.
CC   -!- INDUCTION: By anoxia.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF154418; AAD38397.1; -.
DR   EMBL; AE003688; AAF54549.2; -.
DR   EMBL; AE003688; AAF54550.1; -.
DR   EMBL; AE003688; AAF54551.1; -.
DR   EMBL; AE003688; AAF54552.1; -.
DR   EMBL; AY060997; AAL28545.1; -.
DR   EMBL; AY119569; AAM50223.1; -.
DR   FlyBase; FBgn0020439; fau.
DR   GO; GO:0006979; P:response to oxidative stress; IDA.
KW   Alternative splicing.
FT   DOMAIN       88    143       THR-RICH.
FT   DOMAIN      354    668       ALA/GLU-RICH.
FT   VARSPLIC     26    131       YPSVEKVTRVYKSSYPIYSSYSVPRRVYGATRVVTSPIRVV
FT                                TSPARVVSRVIHSPSPVRVVRTTTRVISSPERTTYSYTTPS
FT                                TYYSPSYLPSTYTSTYIPTSYTTY -> TPRLDLCTDRPGS
FT                                HRSRASSDYSYTSKSSVEKSSYDSSNPHSYRPERSTYTSTV
FT                                EKTSRSGPGGSYNYSTERTSTTGAGPGGYSYSSTTSGNLPG
```

```
FT                                GTKYRHFSYHV (in isoform C).
FT                                /FTId=VSP_004046.
FT   VARSPLIC    132    668       Missing (in isoform C).
FT                                /FTId=VSP_004047.
FT   VARSPLIC     26    163       YPSVEKVTRVYKSSYPIYSSYSVPRRVYGATRVVTSPIRVV
FT                                TSPARVVSRVIHSPSPVRVVRTTTRVISSPERTTYSYTTPS
FT                                TYYSPSYLPSTYTSTYIPTSYTTYTPSYAYSPTTVTRVYAP
FT                                RSSLSPLRITPSPVR -> RTKRTPIDWEKVPFVPRPSLIS
FT                                DPVTAFGVRRPDLERRQRSILDPINRASIKPDYKLAYEPIE
FT                                PYVSTRDKNRTRILGMVRQHIDTVEAGGNTAGRTFRDSLDA
FT                                QLPRLHRAVSESLPVRRETYRNERSGAMVTKYSY (in
FT                                isoform B).
FT                                /FTId=VSP_004048.
FT   VARSPLIC    164    668       Missing (in isoform B).
FT                                /FTId=VSP_004049.
FT   VARSPLIC     36     98       YKSSYPIYSSYSVPRRVYGATRVVTSPIRVVTSPARVVSRV
FT                                IHSPSPVRVVRTTTRVISSPER -> SIWFSSLSPTSLVII
FT                                HLSMDYYYSFLLTYITHCSITSITSSSLEKQIRLVISPHII
FT                                VLFRLKH (in isoform D).
FT                                /FTId=VSP_004050.
FT   VARSPLIC     99    668       Missing (in isoform D).
FT                                /FTId=VSP_004051.
FT   VARSPLIC    280    328       Missing (in isoform E).
FT                                /FTId=VSP_004052.
SQ   SEQUENCE   668 AA;  74487 MW;  4507369593D1DD3B CRC64;

  Query Match             32.1%;  Score 166.5;  DB 1;  Length 668;
  Best Local Similarity   41.0%;  Pred. No. 0.0002;
  Matches   55;  Conservative   16;  Mismatches   30;  Indels   33;  Gaps    4;

Qy          9 EKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK-AKKEAYKAE--A 65
              ||  |:||  || ::  | || :    |||:|:| |:||| |:|||| | :||  ||  |
Db        380 EKKRAQKADEAKRREERALKEERDRLTAEAEKQAAAKAKKAAEEAAKIAAEEALLAEAAA 439

Qy         66 KKYAKAAKA--------------------------EKKEYAAAEAKKAEAAKAYKA-- 95
              :| |: |||                          |  : || ||: || | | ||
Db        440 QKAAEEAKALKAAEDAAQKAAEEARLAEEAAAQKVAEEAAQKAAEEARLAEEAAAQKAAE 499

Qy         96 EAAKAAAKEAAYEA 109
              |||: ||:||| :|
Db        500 EAAQKAAEEAALKA 513


RESULT 9
H110_CHICK
ID   H110_CHICK     STANDARD;      PRT;   219 AA.
AC   P08286;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1.10 (Methylated DNA binding protein-2-H1) (MDBP-2-H1).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87250632; PubMed=3597432;
RA   Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
RT   "Characterization of the chicken histone H1 gene complement.
RT   Generation of a complete set of vertebrate H1 protein sequences.";
RL   J. Biol. Chem. 262:9656-9663(1987).
RN   [2]
RP   IDENTIFICATION OF ALA-13.
RX   MEDLINE=98060905; PubMed=9396815;
RA   Schwarz S., Hess D., Jost J.P.;
RT   "The methylated DNA binding protein-2-H1 (MDBP-2-H1) consists of
RT   histone H1 subtypes which are truncated at the C-terminus.";
RL   Nucleic Acids Res. 25:5052-5056(1997).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17018; AAA48788.1; -.
DR   HSSP; P08287; 1GHC.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN       36    109       GLOBULAR.
SQ   SEQUENCE   219 AA;  21872 MW;  A944DF6C0D3048AE CRC64;

  Query Match             31.8%;  Score 165;  DB 1;  Length 219;
  Best Local Similarity   46.4%;  Pred. No. 0.0001;
  Matches   52;  Conservative   11;  Mismatches   41;  Indels    8;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEA----KAYKAAEAK---KKAKAEAKKYAKEA 53
              :||  :  |||  |:  ||| ||  ||| |    |  ||| ||   ||||  |    |:|
Db        108 SKKPGEVKEKAPRKRTPAAKPKKPAAKKPASAAKKPKKAAAAAKKSPKKAKKPAAAATKKA 167

Qy         54 AKAKKEAYKAEAKKYAKAAKAEKKEYAA-AEAKKAEAAKAYKAEAAKAAAKE 104
              ||: |:| ||   | |  ||:  |  |   :| | :|||   |:| |||||:
Db        168 AKSPKKATKAAKPKKAATAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAAKK 219


RESULT 10
TMPB_TREPH
ID   TMPB_TREPH     STANDARD;      PRT;   384 AA.
AC   P29720;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Treponemal membrane protein B precursor (Antigen tmpB).
GN   TMPB.
OS   Treponema phagedenis.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kazan 5;
RX   MEDLINE=91372983; PubMed=1894368;
RA   Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,
RA   Slivinsky L., Schouls L.M., van Embden J.D., Charon N.W.;
RT   "Treponema phagedenis encodes and expresses homologs of the Treponema
RT   pallidum TmpA and TmpB proteins.";
RL   Infect. Immun. 59:3685-3693(1991).
CC   -!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR
CC       LARGE MOLECULES.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.
CC   -!- SIMILARITY: TO TMPB OF T.PALLIDUM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58563; AAA27480.1; -.
DR   PIR; B43592; B43592.
```

```
KW   Antigen; Outer membrane; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    384       TREPONEMAL MEMBRANE PROTEIN B.
FT   DOMAIN      151    235       17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR]-
FT                                [ED].
FT   REPEAT      151    155       1-1.
FT   REPEAT      156    160       1-2.
FT   REPEAT      161    165       1-3.
FT   REPEAT      166    170       1-4.
FT   REPEAT      171    175       1-5.
FT   REPEAT      176    180       1-6.
FT   REPEAT      181    185       1-7.
FT   REPEAT      186    190       1-8.
FT   REPEAT      191    195       1-9.
FT   REPEAT      196    200       1-10.
FT   REPEAT      201    205       1-11.
FT   REPEAT      206    210       1-12.
FT   REPEAT      211    215       1-13.
FT   REPEAT      216    220       1-14.
FT   REPEAT      221    225       1-15.
FT   REPEAT      226    230       1-16.
FT   REPEAT      231    235       1-17.
FT   DOMAIN      236    288       6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-
FT                                A-A-E.
FT   REPEAT      236    243       2-1.
FT   REPEAT      245    252       2-2.
FT   REPEAT      254    261       2-3.
FT   REPEAT      263    270       2-4.
FT   REPEAT      272    279       2-5.
FT   REPEAT      281    288       2-6.
SQ   SEQUENCE   384 AA;  42677 MW;  6E94CBC74294DE8C CRC64;

  Query Match             31.5%;  Score 163.5;  DB 1;  Length 384;
  Best Local Similarity   46.0%;  Pred. No. 0.0002;
  Matches   52;  Conservative   12;  Mismatches   42;  Indels    7;  Gaps    3;

Qy          2 KKYAK-KAEKAYAKKAKAAKEK----KAYAKKEAKAYKAAEAKKKAKAEAKKYAKE--AA 54
              :| || || |  | | ||||||    ||  :| ||   | |   | ||   | ||| |
Db        180 EKAAKEKAAKDKAAKEKAAKEKAAKDKAAKEKAAKEKAAREMAAKEKAAKDKAAKEEAAR 239

Qy         55 KAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAY 107
              || :||   :| : | | || ::| |   |::  | || : |||: ||:|| |
Db        240 KAAEEAAARKAAEEAAARKAAEEEAARIAAEEEAARKAAEEEAARKAAEEALY 292


RESULT 11
H11L_CHICK
ID   H11L_CHICK     STANDARD;      PRT;   224 AA.
AC   P08287;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Histone H1.11L.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87250632; PubMed=3597432;
RA   Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
RT   "Characterization of the chicken histone H1 gene complement.
RT   Generation of a complete set of vertebrate H1 protein sequences.";
RL   J. Biol. Chem. 262:9656-9663(1987).
RN   [2]
RP   STRUCTURE BY NMR OF 40-113.
RX   MEDLINE=94032251; PubMed=8218199;
RA   Cerf C., Lippens G., Muyldermans S., Segers A., Ramakrishnan V.,
RA   Wodak S.J., Hallenga K., Wyns L.;
RT   "Homo- and heteronuclear two-dimensional NMR studies of the globular
RT   domain of histone H1: sequential assignment and secondary
RT   structure.";
RL   Biochemistry 32:11345-11351(1993).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M17019; AAA48789.1; -.
DR   PIR; B28456; B28456.
DR   PDB; 1GHC; 31-AUG-94.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Acetylation; 3D-structure.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN       40    113       GLOBULAR.
FT   HELIX        44     54
FT   TURN         62     63
FT   HELIX        66     68
FT   HELIX        81     87
FT   TURN         88     90
FT   HELIX        91     94
FT   TURN         95     95
FT   TURN        104    105
SQ   SEQUENCE   224 AA;  22397 MW;  D3D057CB97865CAF CRC64;

  Query Match             31.4%;  Score 163;  DB 1;  Length 224;
  Best Local Similarity   43.8%;  Pred. No. 0.00014;
  Matches   49;  Conservative   13;  Mismatches   42;  Indels    8;  Gaps    2;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKK-------AYAKKEAKAYKAAEAKKKAKAEAKKYAKEA 53
              :||  :  |||  ||| ||| ||        | |||  ||    :: ||||  |    |::
Db        112 SKKPGEVKEKAPKKKASAAKPKKPAAKKPAAAAKKPKKAVAVKKSPKKAKKPAASATKKS 171

Qy         54 AKAKKEAYKAEAKKYAKAAKAEKKEYAA-AEAKKAEAAKAYKAEAAKAAAKE 104
              ||: |:  ||   | | |||:  |  |   :| | :|||   |:| |||||:
Db        172 AKSPKKVTKAVKPKKAVAAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAAKK 223


RESULT 12
H1_ONCMY
ID   H1_ONCMY       STANDARD;      PRT;   206 AA.
AC   P06350;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85264847; PubMed=6443128;
RA   Mezquita J., Connor W., Winkfein R.J., Dixon G.H.;
RT   "An H1 histone gene from rainbow trout (Salmo gairdnerii).";
RL   J. Mol. Evol. 21:209-219(1985).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X02624; CAB37646.1; -.
DR   PIR; A02584; HSTR1R.
DR   HSSP; P08287; 1GHC.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN       27    100       GLOBULAR.
SQ   SEQUENCE   206 AA;  20672 MW;  72C4407980667l6C CRC64;

  Query Match             30.9%;  Score 160.5;  DB 1;  Length 206;
  Best Local Similarity   47.7%;  Pred. No. 0.00019;
  Matches   52;  Conservative    9;  Mismatches   35;  Indels   13;  Gaps    4;

Qy          1 AKKYAK-----KAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              ||| ||     ||:|  |||  |||: |  | |:| | | :  | |  |  || ||   |
Db        105 AKKPAKKAAAPKAKKVAAKKPAAAKKPKKVAAKKAVAAKKSPKKAKKPATPKKAAKSPKK 164

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
                || |  | ||   |||: ||   | :| | :|||   |:| ||| |:
Db        165 VKKPA--AAAK---KAAKSPKK---ATKAAKPKAAKPKAAKAKKAAPKK 205


RESULT 13
H15_HUMAN
ID   H15_HUMAN      STANDARD;      PRT;   225 AA.
AC   P16401; Q14529;
DT   01-AUG-1990 (Rel. 15, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Histone H1.5 (Histone H1a).
GN   HIST1H1B OR H1F5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Spleen;
RX   MEDLINE=90130391; PubMed=2613692;
RA   Ohe Y., Hayashi H., Iwai K.;
RT   "Human spleen histone H1. Isolation and amino acid sequences of three
RT   minor variants, H1a, H1c, and H1d.";
RL   J. Biochem. 106:844-857(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97183654; PubMed=9031620;
RA   Albig W., Meergans T., Doenecke D.;
RT   "Characterization of the H1.5 gene completes the set of human H1
RT   subtype genes.";
RL   Gene 184:141-148(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22296985; PubMed=12408966;
RA   Marzluff W.F., Gongidi P., Woods K.R., Jin J., Maltais L.J.;
RT   "The human and mouse replication-dependent histone genes.";
RL   Genomics 80:487-498(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Wild A.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X83509; CAA58498.1; -.
DR   EMBL; AF531304; AAN06704.1; -.
DR   EMBL; Z98744; CAB11421.1; -.
DR   PIR; S51660; S51660.
DR   HSSP; P08287; 1GHC.
DR   Genew; HGNC:4719; HIST1H1B.
DR   MIM; 142711; -.
DR   GO; GO:0005718; C:nucleosome; NAS.
DR   GO; GO:0003677; F:DNA binding activity; NAS.
DR   GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; NAS.
DR   GO; GO:0006334; P:nucleosome assembly; NAS.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   CONFLICT    215    217       MISSING (IN REF. 1).
SQ   SEQUENCE   225 AA;  22449 MW;  26CD4A1E5D463CDA CRC64;

  Query Match             30.9%;  Score 160.5;  DB 1;  Length 225;
  Best Local Similarity   46.8%;  Pred. No. 0.0002;
  Matches   51;  Conservative    7;  Mismatches   44;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAA-----EAKKKAKAEAKKYAKEAAK 55
              ||  |||| | |||   |  ||  ||| | | ||       :||| | |  || ||   |
Db        118 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKK 175

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              ||  |   :| |      || | : |  :| | :|||   |:| |||||:
Db        176 AKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKK 224


RESULT 14
DBH_MYCSM
ID   DBH_MYCSM      STANDARD;      PRT;   208 AA.
AC   Q9ZHC5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN   HUP OR HLP.
OS   Mycobacterium smegmatis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 700084 / mc(2)155;
RX   MEDLINE=99110209; PubMed=9894918;
RA   Lee B.H., Murugasu-Oei B., Dick T.;
RT   "Upregulation of a histone-like protein in dormant Mycobacterium
RT   smegmatis.";
RL   Mol. Gen. Genet. 260:475-479(1998).
CC   -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
```

```
CC       PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC       DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC       ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068138; AAD13809.1; -.
DR   HSSP; P02346; 1HUU.
DR   InterPro; IPR000119; Bac_DNAbind.
DR   Pfam; PF00216; Bac_DNA_binding; 1.
DR   ProDom; PD000945; Bac_DNAbind; 1.
DR   SMART; SM00411; BHL; 1.
DR   PROSITE; PS00045; HISTONE_LIKE; 1.
KW   DNA-binding; DNA condensation; Repeat.
FT   DOMAIN        1     90       BACTERIAL HISTONE-LIKE DOMAIN.
FT   DOMAIN      101    205       DEGENERATE REPEATS REGION.
SQ   SEQUENCE   208 AA;  21230 MW;  CA5F577F61F7EF09 CRC64;

  Query Match             30.8%;  Score 160;  DB 1;  Length 208;
  Best Local Similarity   54.3%;  Pred. No. 0.00021;
  Matches   57;  Conservative    3;  Mismatches   31;  Indels   14;  Gaps    5;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||| ||||    |||| | |     | |:| | |||      || ||| | : | ||| |
Db        111 AKKAAKKAP---AKKAAAKKTATKAAAKKAPAKKAA-----TKAPAKKAATK-APAKKAA 161

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEA 105
               || ||| |  | | ||  | | ||||    | || | |||||:|
Db        162 TKAPAKKAATKAPA-KKAAAKAPAKKA----ATKAPAKKAAAKKA 201


RESULT 15
H11R_CHICK
ID   H11R_CHICK     STANDARD;      PRT;   218 AA.
AC   P08288;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Histone H1.11R.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87250632; PubMed=3597432;
RA   Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
RT   "Characterization of the chicken histone H1 gene complement.
RT   Generation of a complete set of vertebrate H1 protein sequences.";
RL   J. Biol. Chem. 262:9656-9663(1987).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17020; AAA48790.1; -.
DR   PIR; C28456; C28456.
DR   PDB; 1GHC; 31-AUG-94.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   3D-structure.
FT   INIT_MET      0      0
FT   DOMAIN       37    110       GLOBULAR.
SQ   SEQUENCE   218 AA;  21672 MW;  CB9724BFF14654A6 CRC64;

  Query Match             30.7%;  Score 159.5;  DB 1;  Length 218;
  Best Local Similarity   47.7%;  Pred. No. 0.00023;
  Matches   52;  Conservative   11;  Mismatches   41;  Indels    5;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKK---EAKAYKAAEAKKKAKAEAKKYAKEAAKAK 57
              :|| :  |||  ||| ||| ||| |||    ||  | | | ||:  :||| |  | |
Db        109 SKKPGEGLEKAPKKKASAAKPKKAAAKKPAAAAKKPKKAVAVKKSPKKAKKPAASATKKS 168

Qy         58 KEAYKAEAK-KYAKAAKAEKKEYAA-AEAKKAEAAKAYKAEAAKAAAKE 104
               :: |  || | | |||:  |  |   :| | :|||   |:| |||||:
Db        169 VKSPKKAAKPKKAVAAKSPAKAKAVKPKAAKPKAAKPKAAKAKKAAAKK 217


Search completed: January 21, 2004, 09:00:59
Job time : 15.8636 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:53:31 ; Search time 43.8831 Seconds
(without alignments)
640.969 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	202	38.9	395	2	Q937K4	Q937k4 erwinia chr
2	190.5	36.7	372	2	Q9WWX1	Q9wwx1 pseudomonas
3	190.5	36.7	1866	5	Q8T5C8	Q8t5c8 plasmodium
4	188.5	36.3	421	16	Q8FJT1	Q8fjt1 escherichia
5	184.5	35.5	407	16	Q8ZQT6	Q8zqt6 salmonella
6	181.5	35.0	394	16	Q8X965	Q8x965 escherichia
7	178	34.3	376	16	Q8Z8C1	Q8z8c1 salmonella
8	178	34.3	388	16	Q8ZGZ2	Q8zgz2 yersinia pe
9	178	34.3	393	16	Q8CZZ8	Q8czz8 yersinia pe
10	176.5	34.0	200	16	Q8XVN7	Q8xvn7 ralstonia s
11	173.5	33.4	389	16	Q9CM70	Q9cm70 pasteurella
12	170.5	32.9	1701	5	O61164	O61164 plasmodium
13	170	32.8	232	10	Q39576	Q39576 chlamydomon
14	169.5	32.7	275	5	O01395	O01395 drosophila
15	164	31.6	265	10	Q39598	Q39598 chlamydomon
16	164	31.6	311	12	Q84528	Q84528 paramecium
17	163	31.4	182	2	Q45370	Q45370 bordetella
18	162.5	31.3	1128	5	Q26947	Q26947 trypanosoma
19	162	31.2	356	16	Q9KR10	Q9kr10 vibrio chol
20	161.5	31.1	293	10	Q9AT18	Q9at18 lens culina
21	161.5	31.1	295	10	Q8LKH9	Q8lkh9 pisum fulvu
22	161.5	31.1	295	10	Q8LKI0	Q8lki0 pisum sativ
23	161.5	31.1	295	10	Q9AT22	Q9at22 lathyrus sa
24	161.5	31.1	295	10	Q9ZR20	Q9zr20 pisum sativ
25	161.5	31.1	297	10	Q9SXQ8	Q9sxq8 pisum sativ
26	161.5	31.1	306	10	Q9AT21	Q9at21 lathyrus sa
27	161	31.0	296	10	Q9AT25	Q9at25 pisum sativ
28	160.5	30.9	301	10	Q9AT23	Q9at23 pisum sativ
29	158.5	30.5	298	10	Q8LKI1	Q8lki1 lathyrus ap
30	158	30.4	281	10	Q9AT20	Q9at20 lens culina
31	158	30.4	290	10	Q9AT24	Q9at24 pisum sativ
32	158	30.4	1341	16	Q98KG7	Q98kg7 rhizobium l
33	157.5	30.3	155	16	Q8PI40	Q8pi40 xanthomonas
34	156	30.1	224	13	Q90ZD7	Q90zd7 bufo bufo g
35	156	30.1	885	16	Q8EHL5	Q8ehl5 shewanella
36	154.5	29.8	309	16	Q9HUC3	Q9huc3 pseudomonas
37	154.5	29.8	755	10	Q9FP71	Q9fp71 oryza sativ
38	154	29.7	293	10	Q9AT19	Q9at19 lens culina
39	153.5	29.6	369	16	O83483	O83483 treponema p
40	152	29.3	156	16	Q8P6U0	Q8p6u0 xanthomonas
41	151	29.1	607	5	Q9V6S7	Q9v6s7 drosophila
42	150	28.9	243	5	Q23784	Q23784 chironomus
43	150	28.9	288	10	O65795	O65795 triticum ae
44	148.5	28.6	275	10	Q9XHL9	Q9xhl9 triticum ae
45	148.5	28.6	319	16	Q8DAM4	Q8dam4 vibrio vuln

ALIGNMENTS

RESULT 1
Q937K4
ID Q937K4 PRELIMINARY; PRT; 395 AA.
AC Q937K4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TolA protein.
GN TOLA.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;
RT "characterization of the Erwinia chrysanthemi tol-pal genes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ297885; CAC82708.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZEI.
SQ SEQUENCE 395 AA; 41601 MW; 3C0C1DC12E181013 CRC64;

Query Match 38.9%; Score 202; DB 2; Length 395;
Best Local Similarity 47.3%; Pred. No. 1.1e-06;
Matches 62; Conservative 15; Mismatches 32; Indels 22; Gaps 4;

```
Qy          1 AKKYAKKAEKAYAK-------KAKAAKEKKAYAKKEAKAYKAAEAKKKAK---------- 43
              ||:  |:|| | ||       :|||| : |  |::| |   ||:|||||:
Db        138 AKEQQKQAEVAAAKAKAEAEQQAKAAADAKKQAEEEVKKQAAADAKKKAEEEAKTKAAAA 197

Qy         44 AEAKKYAKEAAKAKKEA---YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYK--AEAA 98
              ||||| |:| |||| |    ||| :  ||||:| |:: ||  ||||||| | |  |:
Db        198 AEAKKKAEEEAKAKAAADAKQKAEEEAKAKAAEAAKEKAAADAAKKAEAAAAAKKAADDK 257

Qy         99 KAAAKEAAYEA 109
              | ||  || :|
Db        258 KKAAATAAKQA 268
```

```
RESULT 2
Q9WWX1
ID   Q9WWX1      PRELIMINARY;      PRT;  372 AA.
AC   Q9WWX1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TolA protein.
GN   TOLA.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RX   MEDLINE=96198174; PubMed=8626299;
RA   Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;
RT   "The Pseudomonas putida peptidoglycan-associated outer membrane
RT   lipoprotein (PAL) is involved in maintenance of the integrity of the
RT   cell envelope.";
RL   J. Bacteriol. 178:1699-1706(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RA   Ramos-Gonzalez I.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RA   Rodriguez-Herva J.J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mt-2;
RX   MEDLINE=96422022; PubMed=8824639;
RA   Rodriguez-Herva J.J., Ramos J.;
RT   "Characterization of an OprL null mutant of Pseudomonas putida.";
RL   J. Bacteriol. 178:5836-5840(1996).
DR   EMBL; X74218; CAB50780.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR006260; TonB_C.
DR   PRINTS; PR00624; HISTONEH5.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
SQ   SEQUENCE   372 AA;  40133 MW;  87F49785ECC3C0BC CRC64;

  Query Match             36.7%;  Score 190.5;  DB 2;  Length 372;
  Best Local Similarity   50.4%;  Pred. No. 6.3e-06;
  Matches   59;  Conservative   14;  Mismatches   31;  Indels   13;  Gaps    4;

Qy          5 AKKAEKAYAKKAKAAK----EKKAYAKKEAKAYKAAE------AKKKAKAEAKKYAKEAA 54
              ||||| | || |:|||    :| | ||| :| ||||      |||||: |||| |:| |
Db        118 AKKAEDA-AKAAEAAKAAEAKKAAEAKKADEAKKAAEKQQADIAKKKAEDEAKKKAEEEA 176

Qy         55 K--AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              |  | :|| |  |:   | |  | |: || :||| ||:  | :||: | |:|| :|
Db        177 KKAAAEEAKKKAAEDAKKKAAEEAKKKAAEDAKKKAAAEDAKKKAAEEAKKKAAADA 233


RESULT 3
Q8T5C8
ID   Q8T5C8      PRELIMINARY;      PRT;  1866 AA.
AC   Q8T5C8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Maebl.
GN   MAEBL.
OS   Plasmodium vivax.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5855;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Salvador;
RA   Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT   "Evolutionary relationships of conserved cysteine-rich motifs in
RT   adhesive molecules of malaria parasites.";
RL   Mol. Biol. Evol. 0:0-0(2002).
DR   EMBL; AY042083; AAL10508.1; -.
SQ   SEQUENCE   1866 AA;  212420 MW;  DC692D7CFAE7D93F CRC64;

  Query Match             36.7%;  Score 190.5;  DB 5;  Length 1866;
  Best Local Similarity   49.6%;  Pred. No. 2.7e-05;
  Matches   61;  Conservative   17;  Mismatches   22;  Indels   23;  Gaps    6;

Qy          5 AKKAEKAYAKKAKAAKE----KKAYAKKEAKAYKAAEAKKKA----------KAEAKKYA 50
              |||||:  |:||:|||:    :|| | |:|:| : ||| |||          |||| : |
Db       1262 AKKAEE--ARKAEAAKKAEEARKAEAAKKAEAARKAEAAKKAEEARKAEEARKAEAARKA 1319

Qy         51 KEAAKAKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAK----AYKAEAAKAAAKEAA 106
              ::| || :|| |||| | |:||:  :    |  |:||||||    | |||||| |  |||
Db       1320 EDARKA-EEARKAEAAKKAEAARKAEAAKKAEAARKAEAAKKAEEARKAEAAKKA--EAA 1376

Qy        107 YEA 109
              :|
Db       1377 RKA 1379


RESULT 4
Q8FJT1
ID   Q8FJT1      PRELIMINARY;      PRT;  421 AA.
AC   Q8FJT1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TolA protein.
GN   TOLA OR C0818.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016757; AAN79291.1; -.
KW   Complete proteome.
SQ   SEQUENCE   421 AA;  43184 MW;  DB296626F056D385 CRC64;

  Query Match             36.3%;  Score 188.5;  DB 16;  Length 421;
  Best Local Similarity   53.0%;  Pred. No. 9.6e-06;
  Matches   61;  Conservative   11;  Mismatches   32;  Indels   11;  Gaps    5;

Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: :|||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADAKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||   |   |||  | |||||:|    |||:  |||| | || | |
Db        178 AAVEAQKKAEAAAAALKKKAEAAEAAAAEARKKAATEAAEKAKAEAEKKAAAEKA 232


RESULT 5
Q8ZQT6
ID   Q8ZQT6      PRELIMINARY;      PRT;  407 AA.
AC   Q8ZQT6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Tol protein, membrane spanning protein.
GN   TOLA OR STM0747.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008730; AAL19691.1; -.
KW   Complete proteome.
SQ   SEQUENCE   407 AA;  41865 MW;  2534352116602D75 CRC64;

  Query Match             35.5%;  Score 184.5;  DB 16;  Length 407;
  Best Local Similarity   55.5%;  Pred. No. 1.7e-05;
  Matches   66;  Conservative   11;  Mismatches   25;  Indels   17;  Gaps    8;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYK-AAEAKKKAKAEAKKYAKEA-AKAKK 58
              ||  |   :|| |: |||| : |   | ||:| | ||:|||||:||| | | :|  ||:
Db        138 AKAAADAKKKAEAEAAKAAADAK--KKAEAEAVKAAADAKKKAEAEAAKAAADAKKKAEA 195

Qy         59 EAYK--AEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAK---AAAKEAAYEA 109
              || |  ||||| |:|  |:    |||||||   ||||||  ||| |    ||| :|| ||
Db        196 EAAKAAAEAKKKAEAEAAK----AAAEAKKKADAEAAKA-AAEAKKKADAAAAKAAAEA 249


RESULT 6
Q8X965
ID   Q8X965      PRELIMINARY;      PRT;   394 AA.
AC   Q8X965;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Membrane spanning protein, required for outer membrane integrity
DE   (Membrane spanning protein TolA).
GN   TOLA OR Z0907 OR ECS0774.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005252; AAG55075.1; -.
DR   EMBL; AP002553; BAB34197.1; -.
DR   InterPro; IPR000104; Antifreeze_1.
DR   PRINTS; PR00308; ANTIFREEZEI.
KW   Complete proteome.
SQ   SEQUENCE   394 AA;  40517 MW;  5B58D8E8230BDE28 CRC64;

  Query Match             35.0%;  Score 181.5;  DB 16;  Length 394;
  Best Local Similarity   52.2%;  Pred. No. 2.7e-05;
  Matches   60;  Conservative   12;  Mismatches   32;  Indels   11;  Gaps    5;

Qy          1 AKKYAKKAE----KAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAK-EAAK 55
              |:: ||:||    :|    |||| : || |: : ||  | || ||| |:||| |: ||||
Db        120 AEEAAKQAELKQKQAEEAAAKAAADAKAKAEADDKA--AEEAAKKAAADAKKKAEAEAAK 177

Qy         56 AKKEAY-KAEAKKYAKAAKAEKKEYAAAEAKK---AEAAKAYKAEAAKAAAKEAA 106
              |  ||  ||||   |   |||  | |||||:|   || | | |  | ||||::||
Db        178 AAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAAAEKAAADKKAAEKAAAEKAA 232


RESULT 7
Q8Z8C1
ID   Q8Z8C1      PRELIMINARY;      PRT;   376 AA.
AC   Q8Z8C1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   TolA protein.
GN   STY0793.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627268; CAD05209.1; -.
KW   Complete proteome.
SQ   SEQUENCE   376 AA;  38804 MW;  EC21F2C4767A8A42 CRC64;

  Query Match             34.3%;  Score 178;  DB 16;  Length 376;
  Best Local Similarity   52.3%;  Pred. No. 4.5e-05;
  Matches   58;  Conservative   10;  Mismatches   33;  Indels   10;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||  |   :|| |: |||| : |  |: || |  ||:|||||:|||   || ||:|||
Db        137 AKAAADAKKKAEAEAAKAAADAKKKAEAEA-AKAAADAKKKAEAEA---AKAAAEAKK- 190

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAA---KEAAYE 108
              ||||:    || |:||  | |    ||| |   | ||||||   |:|| |
Db        191 -KAEAEAAKAAADAKKKADAEAAKAAAEAKKKADAAAAKAAADAKKKAAAE 240


RESULT 8
Q8ZGZ2
ID   Q8ZGZ2      PRELIMINARY;      PRT;   388 AA.
AC   Q8ZGZ2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
DE   TolA colicin import membrane protein.
GN   TOLA OR YPO1123.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414146; CAC89966.1; -.
KW   Complete proteome.
SQ   SEQUENCE   388 AA;  40424 MW;  81447B04B30A7E7C CRC64;

  Query Match             34.3%;  Score 178;  DB 16;  Length 388;
  Best Local Similarity   52.1%;  Pred. No. 4.6e-05;
  Matches   63;  Conservative   12;  Mismatches   28;  Indels   18;  Gaps    7;

Qy          2 KKYAKKAEKAYAKKAKAAKEKK------AYAKKEA-KAYKA-AEAKKKAKAEAKKYAKEA 53
              |: |:: : | |  ||| :|:|      | || || |  || |||:|||:||||| |  |
Db        140 KQAAEQQKIAAAAVAKAKEEQKQAETAAAQAKAEADKIVKAQAEAQKKAEAEAKKEAAVA 199

Qy         54 AKAKKEAYKAEAKKYAKAAK------AEKKEYAAAEAKKAEAAK--AYKAEAAKAAAKEA 105
              | |||:|  |:|||  : |:      ||||  | || ||| |||  |  ||| | || ||
Db        200 AAAKKQA-DADAKKAVEVAEKAAADAAEKKAAADAE-KKAAAAKKVAAAAEAKKKAAAEA 257

Qy        106 A 106
              |
Db        258 A 258


RESULT 9
Q8CZZ8
ID   Q8CZZ8      PRELIMINARY;      PRT;   393 AA.
AC   Q8CZZ8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Membrane spanning protein.
GN   TOLA OR Y3056.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
DR   EMBL; AE013906; AAM86606.1; -.
SQ   SEQUENCE   393 AA;  41012 MW;  1E3E4FF87E533481 CRC64;

  Query Match             34.3%;  Score 178;  DB 16;  Length 393;
  Best Local Similarity   52.1%;  Pred. No. 4.7e-05;
  Matches   63;  Conservative   12;  Mismatches   28;  Indels   18;  Gaps    7;

Qy          2 KKYAKKAEKAYAKKAKAAKEKK------AYAKKEA-KAYKA-AEAKKKAKAEAKKYAKEA 53
              |: |:: : | |  ||| :|:|      | || || |  || |||:|||:||||| |  |
Db        145 KQAAEQQKIAAAAVAKAKEEQKQAETAAAQAKAEADKIVKAQAEAQKKAEAEAKKEAAVA 204

Qy         54 AKAKKEAYKAEAKKYAKAAK------AEKKEYAAAEAKKAEAAK--AYKAEAAKAAAKEA 105
              | |||:|  |:|||  : |:      ||||  | || ||| |||  |  ||| | || ||
Db        205 AAAKKQA-DADAKKAVEVAEKAAADAAEKKAAADAE-KKAAAAKKVAAAAEAKKKAAAEA 262

Qy        106 A 106
              |
Db        263 A 263


RESULT 10
Q8XVN7
ID   Q8XVN7      PRELIMINARY;      PRT;   200 AA.
AC   Q8XVN7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable histone H1 protein.
GN   RSC2793 OR RS00453.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Ralstoniaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646071; CAD16500.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
KW   Complete proteome.
SQ   SEQUENCE   200 AA;  19279 MW;  D3831B590510272D CRC64;

  Query Match             34.0%;  Score 176.5;  DB 16;  Length 200;
  Best Local Similarity   54.5%;  Pred. No. 3.2e-05;
  Matches   67;  Conservative    5;  Mismatches   32;  Indels   19;  Gaps    8;

Qy          1 AKKYAKKAEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAK---KKA----KAEAKKY- 49
              ||| |  |:|  |||| |||:   ||  ||| | | |||  |   |||     ||  ||
Db         42 AKKVA--AKKVAAKKAPAAKKAAVKKVAAKKAAPAKKAAVKKVAAKKAPAAKKAAVKKVA 99

Qy         50 AKEAAKAKKEAYKAEAKKYAKAAK--AEKKEYAAAEAKKAEAAKAYKAE-AAKAAAKEAA 106
              ||:|| ||| | |  | | | |||  | ||    |  |||| |||   |: ||| ||| ||
Db        100 AKKAAPAKKAAVKKVAAKKAPAAKKAAAKK---APAAKKAPAAKKAAAKPAAKPAAKPAA 156

Qy        107 YEA 109
              :|
Db        157 KKA 159


RESULT 11
Q9CM70
ID   Q9CM70      PRELIMINARY;      PRT;   389 AA.
AC   Q9CM70;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   TolA.
GN   TOLA OR PM0968.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006136; AAK03052.1; -.
DR   HSSP; P19934; 1TOL.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR00194; TROPOMYOSIN.
KW   Complete proteome.
SQ   SEQUENCE   389 AA;  42197 MW;  B4032F2A2FD9E94B CRC64;

  Query Match             33.4%;  Score 173.5;  DB 16;  Length 389;
  Best Local Similarity   48.7%;  Pred. No. 9.4e-05;
  Matches   55;  Conservative   18;  Mismatches   31;  Indels    9;  Gaps    4;

Qy          6 KKAEKAYAKK-AKAAKEK-KAYAKKEAKAYKAAEAKKKAKA--EAKKYAKEA-----AKA 56
              |:||:| ||: |:||| | :| ||: |   | || : |||| |||: |::|     |||
Db        145 KQAEEAKAKQLAEAAKLKAEAEAKRLAALAKQAEEEAKAKAAEEAKRKAEKAKAEAEAKA 204

Qy         57 KKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              | |  ||||:  || ||: :  | |:|:||:|    ||| ||| |: |  :|
Db        205 KVEKAKAEAEAKVKAEKAKAEAEAKAKAEKAKAEAKAKAEKAKADAEAAQRKA 257


RESULT 12
O61164
ID   O61164      PRELIMINARY;      PRT;  1701 AA.
AC   O61164;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Erythrocyte binding protein.
GN   MAEBL.
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YM;
RX   MEDLINE=98115903; PubMed=9448314;
RA   Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT   "A family of chimeric erythrocyte binding proteins of malaria
RT   parasites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
DR   EMBL; AF031886; AAC05366.1; -.
SQ   SEQUENCE   1701 AA;  199268 MW;  EDA8E2DEFD87CE8A CRC64;

  Query Match             32.9%;  Score 170.5;  DB 5;  Length 1701;
  Best Local Similarity   53.5%;  Pred. No. 0.00058;
  Matches   54;  Conservative   11;  Mismatches   31;  Indels    5;  Gaps    5;

Qy          5 AKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKA- 63
              |||||:   |||:| | |   |||:|:| | || :|| |||| | | |  |  : | ||
Db       1205 AKKAEEE-RKKAEAVK-KAEEAKKKAEAAKKAEERKK-KAEAAKKALERKKKSEAAKKAL 1261

Qy         64 EAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              | || |:|||  :::  |  |||||  |  |||||| | :|
Db       1262 ERKKKAEAAKKAEEKKKAEAAKKAEEEKK-KAEAAKKAEEE 1301


RESULT 13
Q39576
ID   Q39576      PRELIMINARY;      PRT;  232 AA.
AC   Q39576;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Histone H1.
GN   CH1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96120862; PubMed=8590479;
RA   Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R.;
RT   "The organization structure and regulatory elements of Chlamydomonas
RT   histone genes reveal features linking plant and animal genes.";
RL   Curr. Genet. 28:333-345(1995).
DR   EMBL; U16726; AAA98452.1; -.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
SQ   SEQUENCE   232 AA;  24693 MW;  2D006AE44A8FA037 CRC64;

  Query Match             32.8%;  Score 170;  DB 10;  Length 232;
  Best Local Similarity   46.3%;  Pred. No. 0.0001;
  Matches   57;  Conservative   10;  Mismatches   38;  Indels   18;  Gaps    5;

Qy          2 KKYAKKA---EKAYAKKAKAAKEKKAYAKKEA----KAYKAAEAKKKAKAEAKKYAKEA- 53
              || ||||   :|| | | : | :|    |||      |: | || | | : | || || |
Db        100 KKAAKKAATPKKAAAPKKEGAVKKTKAPKKEGEKKPKSAKKAEKKPKKEGEKKKAAKPAK 159

Qy         54 ----AKAKKEAYKAEAKKYAKAAKAEKKEYAAAE---AKKAEAAK---AYKAEAAKAAAK 103
                  || | |   || |   ||||| |||:    || :|||     ||:||| |||
Db        160 AEKKPKAAKPAKKTPTKKAAAKPKAEKKPKAAAKPKAEKKPKAAKPKAEKKAKAAKPAAK 219

Qy        104 EAA 106
              ::|
Db        220 KSA 222


RESULT 14
O01395
ID   O01395      PRELIMINARY;      PRT;  275 AA.
AC   O01395;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Axoneme-associated protein MST101(3).
GN   MST101(3) OR DHMST101.
OS   Drosophila hydei (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7224;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Neesen J., Heinlein U.A.O., Buenemann H.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC   -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR   EMBL; U85627; AAB51369.1; -.
DR   FlyBase; FBgn0020732; Dhyd\mst101(3).
KW   Sperm; Repeat; Multigene family.
FT   DOMAIN       64    255       13 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT                                X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-X-
FT                                [AE]-X.
SQ   SEQUENCE   275 AA;  30436 MW;  76BAA7B2A2DF732C CRC64;

  Query Match             32.7%;  Score 169.5;  DB 5;  Length 275;
  Best Local Similarity   48.2%;  Pred. No. 0.00013;
  Matches   54;  Conservative   16;  Mismatches   29;  Indels   13;  Gaps    6;

Qy          1 AKKYAKKAEKAYAKK--AKAAKEKKAYAKKEAKAYKAAEAKKKAK-AEAKKYAKEAAKAK 57
```

```
              |:   || |:| |||  |:|||::|  |:|:    | ||| || | |  ||   |||| :
Db         51 AEDVKKKCEEAAAKKKCAEAAKKEKEAAEKK----KCAEAAKKEKEAAEKKKCAEAAKKE 106

Qy         58 KEAYKAEAKKYAKAAKAEKKEYAAAEAKK-AEAAKAYKAEAAKAAAKEAAYE 108
              :||  |: || |: || ||:   ||| || |||||  |  | :   :|||::
Db        107 QEA--AQKKKCAELAKKEKE---AAEKKKCAEAAKKEKEAAERKKCEEAAFK 153


RESULT 15
Q39598
ID   Q39598      PRELIMINARY;      PRT;   265 AA.
AC   Q39598;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cgcr-4 product (Fragment).
GN   CGCR-4.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92119224; PubMed=1731966;
RA   Wakarchuk W.W., Muller F.W., Beck C.F.;
RT   "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT   arrangements of directly repeated sequence motifs.";
RL   Plant Mol. Biol. 18:143-146(1992).
DR   EMBL; X17208; CAA35080.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   265 AA;  26216 MW;  B35318B7377CF782 CRC64;

  Query Match             31.6%;  Score 164;  DB 10;  Length 265;
  Best Local Similarity   39.4%;  Pred. No. 0.00029;
  Matches   50;  Conservative   21;  Mismatches   38;  Indels   18;  Gaps    2;

Qy          1 AKKYAKKAEKA--------------YAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEA 46
              |:  ||:|::|               | :|:|  | :|:|  ||:|: |||  : :| |:|
Db         79 AEAKAKEADEARAAEAEARAVAEAWAAAEAEARAEAEAWAAAEAEAWAAAEVEARAAAQA 138

Qy         47 KKYAKEAAKAKKEAYKAEAKKYAKAAK----AEKKEYAAAEAKKAEAAKAYKAEAAKAAA 102
              :  |:  | |  ||  | | : | |||    || :  |||||: | ||:|  |   |:|||
Db        139 RVAAEARAAAAAEARAAAAAEAAAAAKARVAAEARAAAAAEARAAAAAEAAAAAKARAAA 198

Qy        103 KEAAYEA 109
              :  | ||
Db        199 EAKAKEA 205


Search completed: January 21, 2004, 09:00:27
Job time : 44.8831 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:52:56 ; Search time 52.3766 Seconds
(without alignments)
330.323 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	519	100.0	109	21	AAY82577	Copolymer molecula
2	312.5	60.2	86	21	AAY82576	Copolymer molecula
3	289	55.7	77	21	AAY82575	Copolymer molecula
4	228.5	44.0	66	21	AAY82574	Copolymer molecula
5	209	40.3	154	11	AAR06445	Recombinant copoly
6	187	36.0	106	11	AAR06446	Recombinant copoly
7	180.5	34.8	56	21	AAY82573	Copolymer molecula
8	179.5	34.6	100	21	AAY98499	Peptide #10 used i
9	179.5	34.6	100	21	AAY59044	Amino acid polymer
10	179.5	34.6	100	22	AAU04289	Poly-Lys-Ala used
11	179.5	34.6	100	22	AAB45852	Nucleic acid trans
12	179	34.5	347	24	ABJ18771	Pseudomonas aerugi
13	178	34.3	223	20	AAY14928	Amino acid sequenc
14	162.5	31.3	643	16	AAR84568	Trypanosoma cruzi
15	160.5	30.9	158	20	AAY34068	Histone H1 isoform
16	160.5	30.9	158	21	AAY57366	Human histone H1.5
17	160.5	30.9	226	20	AAY34060	Human histone H1 i
18	160.5	30.9	226	21	AAY57358	Human histone H1 i
19	159.5	30.7	234	23	AAU09944	Histone H1-4, frag
20	159.5	30.7	234	23	AAE13234	Human linker histo
21	158.5	30.5	472	16	AAR84569	Trypanosoma cruzi
22	158	30.4	218	18	AAW29477	Human histone H1 i
23	158	30.4	218	20	AAY34034	Human histone H1 i
24	158	30.4	218	21	AAY57332	Human histone H1 i
25	157.5	30.3	564	16	AAR84565	Trypanosoma cruzi
26	154.5	29.8	130	23	AAU09945	Peptide fragment o
27	154.5	29.8	130	23	AAE13237	Human linker histo
28	154	29.7	219	21	AAY76981	Quail H1 histone p
29	153.5	29.6	372	23	ABG80418	Haemophilus influe
30	152.5	29.4	222	18	AAW29476	Human histone H1 i
31	152.5	29.4	222	20	AAY34033	Human histone H1 i
32	152.5	29.4	222	21	AAY57331	Human histone H1 i
33	151	29.1	607	22	ABB62173	Drosophila melanog
34	150	28.9	1507	21	AAB24128	Plasmodium yoelii
35	148.5	28.6	140	13	AAR25206	Synthetic helical
36	146.5	28.2	80	23	ABG71044	Tumour neoangiogen
37	146	28.1	299	22	ABB63276	Drosophila melanog
38	144	27.7	111	18	AAW44829	Adhesive polypepti
39	144	27.7	434	18	AAW21591	Antibiotic potenti
40	144	27.7	434	21	AAY56903	(Lys)434 protein s
41	143.5	27.6	452	24	ABP57088	Mouse MKP4 protein
42	143	27.6	214	20	AAY34055	M. tuberculosis hi
43	143	27.6	214	21	AAY57353	M. tuberculosis hi
44	141.5	27.3	239	23	ABB49123	Listeria monocytog
45	141	27.2	100	21	AAY98497	Peptide #8 used in

ALIGNMENTS

RESULT 1
AAY82577
ID AAY82577 standard; peptide; 109 AA.
XX
AC AAY82577;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 109 AA;

Query Match 100.0%; Score 519; DB 21; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.5e-36;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60

Qy 61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109

RESULT 2
AAY82576
ID AAY82576 standard; peptide; 86 AA.
XX
AC AAY82576;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX

PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 86 AA;

Query Match 60.2%; Score 312.5; DB 21; Length 86;
Best Local Similarity 72.1%; Pred. No. 2.2e-19;
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5;

Qy 1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
||||||| |||||||| :|| | ||||||||||||||
Db 1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA---------- 36

Qy 61 YKAEAKKYAKAAKAEKKEYAAAEAK-KAEAA-KAYKAEAAKAAAKEAAYEA 109
|||||||||||||||||||||||| ||||| ||||||||||||||||||
Db 37 -KAEAKKYAKAAKAEKKEYAAAEAKYKAEAAKKAYKAEAAKAAAKEAAYEA 86

RESULT 3
AAY82575
ID AAY82575 standard; peptide; 77 AA.
XX
AC AAY82575;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;

KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 77 AA;

Query Match 55.7%; Score 289; DB 21; Length 77;
Best Local Similarity 67.0%; Pred. No. 1.8e-17;
Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4;

```
Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA----------------- 36

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              |||||||||||||||||||||||       ||||||||||||||||
Db         37 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 77
```

RESULT 4
AAY82574
ID AAY82574 standard; peptide; 66 AA.
XX
AC AAY82574;
XX
DT 28-JUL-2000 (first entry)
XX
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.
XX
KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX
OS Unidentified.
XX
PN WO200018794-A1.
XX
PD 06-APR-2000.
XX
PF 24-SEP-1999; 99WO-US22402.
XX
PR 25-SEP-1998; 98US-0101693.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
PI Gad A, Lis D;
XX
DR WPI; 2000-317499/27.
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.
XX
CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
SQ Sequence 66 AA;

Query Match 44.0%; Score 228.5; DB 21; Length 66;
Best Local Similarity 56.9%; Pred. No. 1.8e-12;
Matches 62; Conservative 0; Mismatches 4; Indels 43; Gaps 4;

```
Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||  ||          |||| | |
Db          1 AKKYAKK-EKAYAKAKKA----------EAKAAKKA----------------------- 25

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
```

```
                ||||||||||||||||||||||||      ||||||||||||||||||
Db         26 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 66


RESULT 5
AAR06445
ID   AAR06445 standard; protein; 154 AA.
XX
AC   AAR06445;
XX
DT   25-MAR-2003  (updated)
DT   03-JAN-1991  (first entry)
XX
DE   Recombinant copolymer 1-77, myelin basic protein analogue.
XX
KW   Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;
KW   immunological activity; autoimmune encephalomyelitis;
KW   multiple sclerosis;
XX
OS   Synthetic.
XX
PN   EP383620-A.
XX
PD   22-AUG-1990.
XX
PF   16-FEB-1990;   90EP-0301700.
XX
PR   07-FEB-1990;   90US-0473845.
PR   17-FEB-1989;   89US-0312541.
XX
PA   (REPK ) REPLIGEN CORP.
XX
PI   Cook KS;
XX
DR   WPI; 1990-255848/34.
DR   N-PSDB; AAQ05664.
XX
PT   Producing genes encoding random polymers of aminoacid(s) - for
PT   producing recombinant polypeptide(s) with biological and/or
PT   immunological activity
XX
PS   Disclosure; Fig 11; 25pp; English.
XX
CC   To improve the expression of rCOP-1 polypeptides in E. coli, genes
CC   coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN
CC   (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to
CC   express Protein A.  The resulting plasmids encode fusion proteins
CC   consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.
CC   A methionine residue occurs between the Protein A and rCOP-1
CC   sequences, originating from the 5' linker sequence, in order that
CC   the COP-1 polypeptide may be cleaved from the fusion protein.
CC   rCOP-1-77 contains oligonucleotide duplexes incoding the following
CC   segments: YKK, EAE, KAK, AAK, and AAA.  The N-terminal alanine residue
CC   is left behind following CNBr cleavage of the fusion protein.
CC   The product prevents or arrests experimental autoimmune
CC   encephalomyelitis.  They are used to prevent, arrest or control a
CC   demyelinating disorder, e.g. multiple sclerosis.  They may also
CC   be used as additives  to hair care products to confer beneficial
CC   effects on damaged hair or as supplements for diets deficient in
CC   certain amino acids.
CC   See also AAQ05665.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   154 AA;

  Query Match             40.3%;  Score 209;  DB 11;  Length 154;
  Best Local Similarity   53.7%;  Pred. No. 1.9e-10;
  Matches   72;  Conservative    8;  Mismatches   20;  Indels   34;  Gaps    9;

Qy          2 KKYAKKAEKA-------YAKKAKAAKEKKAYA----KKEAKAYKAAEA-----KKKAK-A 44
              || ||:||||       | |:|:|||  || |    ||||:|  ||||     ||||| |
Db         24 KKKAKEAEKAKKAKYKKYKKEAEAAKAAKAAAAAAAYKKEAEA--AAEAEKAKYKKKAKEA 81

Qy         45 EAKKYAKEAAKA--KKEAYKAEAKKYAKAAKAEK---------KEYAAAEAKKAEAA--K 91
              | || || ||:|  |||| :|| ||| | ||  |          |  |||||:||: |  |
Db         82 EYKKKAKAAAEAEYKKEAEEAEYKKYKKKAKKAKYKKKAKEAEKAKAAAEAEKAKEAEYK 141

Qy         92 AYKAEAAKAAAKEA 105
              || || |  ||||
Db        142 KYKKEAEK--AKEA 153


RESULT 6
AAR06446
ID   AAR06446 standard; protein; 106 AA.
XX
AC   AAR06446;
XX
DT   25-MAR-2003  (updated)
DT   03-JAN-1991  (first entry)
XX
DE   Recombinant copolymer 1-19, myelin basic protein analogue.
XX
KW   Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;
KW   immunological activity; autoimmune encephalomyelitis;
KW   multiple sclerosis;
XX
OS   Synthetic.
XX
PN   EP383620-A.
XX
PD   22-AUG-1990.
XX
PF   16-FEB-1990;   90EP-0301700.
XX
PR   07-FEB-1990;   90US-0473845.
PR   17-FEB-1989;   89US-0312541.
XX
PA   (REPK ) REPLIGEN CORP.
XX
PI   Cook KS;
XX
DR   WPI; 1990-255848/34.
DR   N-PSDB; AAQ06446.
XX
PT   Producing genes encoding random polymers of aminoacid(s) - for
PT   producing recombinant polypeptide(s) with biological and/or
PT   immunological activity
XX
PS   Disclosure; Fig 12; 25pp; English.
XX
CC   To improve the expression of rCOP-1 polypeptides in E. coli, genes
CC   coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN
CC   (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to
CC   express Protein A.  The resulting plasmids encode fusion proteins
CC   consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.
CC   A methionine residue occurs between the Protein A and rCOP-1
CC   sequences, originating from the 5' linker sequence, in order that
CC   the COP-1 polypeptide may be cleaved from the fusion protein.
CC   rCOP-1-19 contains oligonucleotide duplexes incoding the following
CC   segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA.  The
CC   N-terminal alanine residue is left behind following CNBr cleavage of the
CC   fusion protein.
CC   The product prevents or arrests experimental autoimmune
CC   encephalomyelitis.  They are used to prevent, arrest or control a
CC   demyelinating disorder, e.g. multiple sclerosis.  They may also
CC   be used as additives  to hair care products to confer beneficial
CC   effects on damaged hair or as supplements for diets deficient in
CC   certain amino acids.
CC   See also AAQ05664.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   106 AA;

  Query Match             36.0%;  Score 187;  DB 11;  Length 106;
  Best Local Similarity   55.0%;  Pred. No. 8.7e-09;
```

```
  Matches   60;  Conservative   11;  Mismatches   20;  Indels   18;  Gaps    6;

Qy          5 AKKAAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              || |||| | | || : :|| || |||  || :|:|  || |:| |||| ||:|
Db          1 AKAAEKAKAAK-KAYEAEKAKAKYEAK--KAEKAEKAEKAAAEKKAKEAKKAEK------ 51

Qy         65 AKKYAKAAKAEKKEYAAAEAKKAEAA----KAYKAEAAKAAAK-EAAYE 108
              ||:  || :|||    | |||||:||    || ||| |||| | :|||:
Db         52 AKEAEKAKEAEK----AKEAKKAKAAEKAEKAKKAEKAKAAEKAKAAYK 96


RESULT 7
AAY82573
ID   AAY82573 standard; peptide; 56 AA.
XX
AC   AAY82573;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.
XX
KW   Copolymer; molecular weight marker; TV-marker; immune disease;
KW   glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW   osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW   antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW   antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW   inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW   Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW   diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW   Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW   pemphigus vulgaris; systemic lupus erythematosus.
XX
OS   Unidentified.
XX
PN   WO200018794-A1.
XX
PD   06-APR-2000.
XX
PF   24-SEP-1999;   99WO-US22402.
XX
PR   25-SEP-1998;   98US-0101693.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (TEVA-) TEVA PHARM USA INC.
XX
PI   Gad A,  Lis D;
XX
DR   WPI; 2000-317499/27.
XX
PT   Copolymer 1 related polypeptides used as molecular weight markers for
PT   glatiramer acetate and for treatment and prevention of immune diseases
XX
PS   Claim 10; Page 14; 72pp; English.
XX
CC   AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC   weight TV-marker polypeptides from the present invention. The present
CC   invention describes polypeptides (I) for determining the molecular
CC   weight of a copolymer (CP), which has an identified molecular weight
CC   and an amino acid composition corresponding to the copolymer. The
CC   polypeptides of the invention are used as molecular weight markers for
CC   glatiramer acetate related tetrapolymers. The polypeptides may also be
CC   used for treating and preventing immune diseases in a mammal. Autoimmune
CC   diseases which may be treated include either cell-mediated or
CC   antibody-mediated diseases. Such diseases include arthritic conditions,
CC   demyelinating diseases and inflammatory conditions, e.g. multiple
CC   sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC   anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC   uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC   purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC   disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC   myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC   lupus erythematosus. Mediated-mediated diseases which can be treated
CC   include host-versus-graft disease, graft-versus-host disease, and
CC   delayed-type hypersensitivity. The polypeptides of the invention have
CC   defined molecular weights and physical properties which are analogous to
CC   glatiramer acetate molecules, which makes them ideal for use as
CC   molecular weight markers.
XX
SQ   Sequence   56 AA;

  Query Match             34.8%;  Score 180.5;  DB 21;  Length 56;
  Best Local Similarity   45.9%;  Pred. No. 1.6e-08;
  Matches   50;  Conservative    3;  Mismatches    3;  Indels   53;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | |||||||||||||||:|:
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKAEAK--------------- 39

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
                                                ||||||||||||||||
Db         40 --------------------------------YKAEAAKAAAKEAAYEA 56


RESULT 8
AAY98499
ID   AAY98499 standard; Peptide; 100 AA.
XX
AC   AAY98499;
XX
DT   31-JUL-2000  (first entry)
XX
DE   Peptide #10 used in nucleic acid transporter system.
XX
KW   Transporter system; nucleic acid delivery; gene therapy; cancer;
KW   carcinogenesis; cardiovascular disease; infection.
XX
OS   Synthetic.
XX
PN   US6033884-A.
XX
PD   07-MAR-2000.
XX
PF   14-DEC-1993;   93US-0167641.
XX
PR   20-MAR-1992;   92US-0855389.
PR   19-MAR-1993;   93WO-US02725.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Gottchalk S,  Sparrow J,  Cristiano RJ,  Woo SLC,  Smith LC;
XX
DR   WPI; 2000-281993/24.
XX
PT   System for transporting nucleic acid into cells, useful e.g. in gene
PT   therapy and for generating transgenic animals, comprises binding agent
PT   linked to nucleic acid, surface ligand and lytic agent  -
XX
PS   Disclosure; Column 125-128; 108pp; English.
XX
CC   The present invention relates to a transporter system for delivering
CC   nucleic acid to a cell. The system comprises a nucleic acid binding
CC   complex, consisting of a binding molecule bonded non-covalently to the
CC   nucleic acid, and covalently to a surface ligand, and a lytic agent. The
CC   binding molecule is spermine or a spermidine derivative. Nucleotide
CC   sequences AAA36633-A36652 and peptide sequences AAY98456-Y98500 are used
CC   in the construction of the transporter system of the invention. The
CC   transporter system is used in gene therapy, particularly to deliver
CC   nucleic acids to hepatocytes, muscle cells or bone forming cells, e.g
CC   for treating cardiovascular disease, cancer, and infection. The
CC   transporter systems are also used to create transgenic animals (as models
CC   for human carcinogenesis or disease or for drug testing). Other uses
CC   include transforming cells to produce proteins , or transfecting cells in
CC   vitro to study the function of the nucleic acid. The use of a surface
CC   ligand allows specific targeting of selected cells and tissues. The lytic
CC   agent provides for release of the nucleic acid into the cellular
CC   interior, from endosomes, without requiring endosomal or lysosomal
```

```
CC   degradation.
XX
SQ   Sequence   100 AA;

  Query Match             34.6%;  Score 179.5;  DB 21;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 3.5e-08;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || |||| : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              || ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 9
AAY59044
ID   AAY59044 standard; peptide; 100 AA.
XX
AC   AAY59044;
XX
DT   07-MAR-2000  (first entry)
XX
DE   Amino acid polymer seq ID NO: 64 of US5994109.
XX
KW   Nucleic acid transport system; NTS; cell surface receptor; cytosis;
KW   nuclear membrane; lysis moiety; transgenic animal; human disease;
KW   nucleic acid delivery; cancer.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3..100
FT                   /note= "Lys-Ala in positions 3 to 100 may be optionally
FT                          absent"
XX
PN   US5994109-A.
XX
PD   30-NOV-1999.
XX
PF   03-JUN-1995;   95US-0460890.
XX
PR   14-DEC-1993;   93US-0167641.
PR   20-MAR-1992;   92US-0855389.
PR   19-MAR-1993;   93WO-US02725.
PR   14-DEC-1993;   93US-0167641.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Woo SLC,  Cristiano RJ,  Gottchalk S,  Sparrow J,  Smith LC;
XX
DR   WPI; 2000-038262/03.
XX
PT   Nucleic acid transport system, useful for creating transgenic animals
PT   for assessing human disease such as cancer in an animal model -
XX
PS   Disclosure; Columns 123-124; 107pp; English.
XX
CC   The invention relates to a nucleic acid transport system (NTS) for
CC   delivering nucleic acid into a cell. The NTS contains but is not limited
CC   to 5 components: (a) the nucleic acid or a macromolecule to be delivered;
CC   (b) a moiety that recognizes and binds to a cell surface receptor or
CC   antigen or is capable of entering a cell through cytosis; (c) a nucleic
CC   acid or macromolecular molecule binding moiety; (d) a moiety that is
CC   capable of moving or initiating movement through a nuclear membrane; and/
CC   or (e) a lysis moiety that enables the transport of the entire complex
CC   from the cell surface directly into the cytoplasm of the cell. The NTS
CC   delivers nucleic acid into the cellular interior as well as the nucleus
CC   of specific cells. The NTS can be used to treat disorders by targeting
CC   specific nucleic acid accordingly. The NTS can also be used to create
CC   transgenic animals for assessing human disease, such as cancer, in an
CC   animal model. The NTS can be used in vitro with tissue culture cells
CC   which allows the role of various nucleic acids to be studied by targeting
CC   specific expression into specifically targeted tissue culture cells. The
CC   lysis agent within the NTS avoids the problem of endosomal/lysosomal
CC   degradation.
XX
SQ   Sequence   100 AA;

  Query Match             34.6%;  Score 179.5;  DB 21;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 3.5e-08;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || |||| : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              || ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 10
AAU04289
ID   AAU04289 standard; Peptide; 100 AA.
XX
AC   AAU04289;
XX
DT   23-OCT-2001  (first entry)
XX
DE   Poly-Lys-Ala used in nucleic acid transporter system.
XX
KW   Nucleic acid transport; cytosis; ligand; lysis agent; spacer molecule;
KW   gene therapy; hepatocyte; muscle; bone forming cell.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3..100
FT                   /note= "Lys-Ala in positions 3-100 may be present or
FT                          absent"
XX
PN   US6177554-B1.
XX
PD   23-JAN-2001.
XX
PF   05-JUN-1995;   95US-0462040.
XX
PR   14-DEC-1993;   93US-0167641.
PR   20-MAR-1992;   92US-0855389.
PR   19-MAR-1993;   93WO-US02725.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Woo SLC,  Smith LC,  Cristiano RJ,  Gottchalk S,  Sparrow J;
XX
DR   WPI; 2001-365933/38.
XX
PT   Nucleic acid transport system, useful for creating transgenic animals
PT   for assessing human disease such as cancer in an animal model -
XX
PS   Disclosure; Column 131; 111pp; English.
XX
CC   The sequence represents poly-Lys-Ala, used to bind nucleic acid in a
CC   nucleic acid transporter system. The nucleic acid transporter system uses
CC   nucleic acid binding complexes containing surface ligands which are
CC   capable of binding to a cell surface receptor and entering the cell
CC   through cytosis. The compounds of the invention are either ligands,
CC   binding molecules (surface ligands), lysis agents, spacer molecules or
CC   their intermediates. The ligands, binding molecules, lysis agents and
CC   spacer molecules are used in nucleic acid transporter systems to deliver
CC   nucleic acid into specific cells e.g. in gene therapy to deliver nucleic
CC   acid into hepatocytes, muscle cells or bone forming cells.
XX
```

```
SQ   Sequence   100 AA;

  Query Match            34.6%;  Score 179.5;  DB 22;  Length 100;
  Best Local Similarity  55.0%;  Pred. No. 3.5e-08;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
                ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 11
AAB45852
ID   AAB45852 standard; Protein; 100 AA.
XX
AC   AAB45852;
XX
DT   21-MAR-2001  (first entry)
XX
DE   Nucleic acid transporter system peptide ligand SEQ ID NO 64.
XX
KW   Nucleic acid delivery; nucleic acid transporter system; hormone; enzyme;
KW   growth factor; clotting factor; apolipoprotein; receptor; drug; oncogene;
KW   tumor antigen; tumor suppressor; viral antigen; parasitic antigen;
KW   bacterial antigen.
XX
OS   Unidentified.
XX
PN   US6150168-A.
XX
PD   21-NOV-2000.
XX
PF   05-JUN-1995;   95US-0460971.
XX
PR   14-DEC-1993;   93US-0167641.
PR   20-MAR-1992;   92US-0855389.
PR   19-MAR-1993;   93WO-US02725.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Gottchalk S,  Sparrow J,  Cristiano RJ,  Smith LC,  Woo SLC;
XX
DR   WPI; 2001-049093/06.
XX
PT   Nucleic acid transporter system for delivering nucleic acid into a
PT   cell, useful for delivering proteins and polypeptides to cells,
PT   including growth factors, enzymes, hormones, and tumor suppressors  -
XX
PS   Disclosure; Column 125-126; 105pp; English.
XX
CC   This invention describes a novel system (I) for delivering a nucleic acid
CC   to a cell, comprising a binding complex comprising a ligand binding
CC   molecule noncovalently bound to a nucleic acid and covalently linked to a
CC   surface ligand, and a second binding complex comprising a second binding
CC   molecule noncovalently bound to a nucleic acid and covalently linked to a
CC   nuclear ligand. The complexes are simultaneously bound to the nucleic
CC   acid. The nucleic acid transporter system can also be used in a method
CC   for the in vivo targeting of the insertion of DNA into a cell. It can
CC   also be used in processes for producing transformed cell lines. The
CC   system can be used to deliver a variety or proteins and polypeptides,
CC   such as hormones, growth factors, enzymes, clotting factors,
CC   apolipoproteins, receptors, drugs, oncogenes, tumor antigens, tumor
CC   suppressors, viral antigens, parasitic antigens, and bacterial antigens.
CC   The transporter system uses lysis agents to overcome the problems of
CC   endosomal/lysosomal degradation seen with prior art systems.
XX
SQ   Sequence   100 AA;

  Query Match            34.6%;  Score 179.5;  DB 22;  Length 100;
  Best Local Similarity  55.0%;  Pred. No. 3.5e-08;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
                ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 12
ABJ18771
ID   ABJ18771 standard; Protein; 347 AA.
XX
AC   ABJ18771;
XX
DT   27-FEB-2003  (first entry)
XX
DE   Pseudomonas aeruginosa biofilm formation-related protein #35.
XX
KW   Biofilm formation modulation; biofilm-associated disease;
KW   cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
KW   catheter-associated infection; medical device-associated infection.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200285295-A2.
XX
PD   31-OCT-2002.
XX
PF   19-APR-2002; 2002WO-US12532.
XX
PR   20-APR-2001; 2001US-285190P.
PR   24-OCT-2001; 2001US-344142P.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Whiteley M,  Bangera MG,  Lory S,  Greenberg EP;
XX
DR   WPI; 2003-075601/07.
DR   N-PSDB; ABT14593.
XX
PT   Identifying compound capable of modulating biofilm formation by
PT   bacteria/bacterial antibiotic resistance, useful for treatment of
PT   biofilm associated disease -
XX
PS   Claim 1; Page 119-120; 154pp; English.
XX
CC   The invention comprises a method for identifying a compound capable of
CC   modulating biofilm formation by bacteria. The method of the invention is
CC   useful for identifying a compound capable of modulating biofilm formation
CC   by bacteria or modulating bacterial antibiotic resistance. The method of
CC   the invention is also useful for diagnosing and treating a subject
CC   (especially an immunocompromised human) that is afflicted with a biofilm-
CC   associated disease or disorder, such as: cystic fibrosis; AIDS; middle
CC   ear infections; acne; periodontal disease; catheter-associated
CC   infections; and medical device-associated infections. The present amino
CC   acid sequence represents a protein that is used in the invention.
XX
SQ   Sequence   347 AA;

  Query Match            34.5%;  Score 179;  DB 24;  Length 347;
  Best Local Similarity  49.1%;  Pred. No. 1.4e-07;
  Matches   54;  Conservative   17;  Mismatches   29;  Indels   10;  Gaps    4;

Qy          6 KKAEKAYAKKAKAAKEK------KAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              :| |:     ||||::|       || |:| |:| || |||| |:|:| :  |:|  |||
Db         99 QKLEQQQVAAAKAAEQKKADEARKAEAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKR 158

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
```

```
                | : ||||  |||:  ||: || :|||  | :| |  ||:|| |:|| ||
Db        159 A-EDEAKK--KAAEDAKKK-AAEDAKKKAAEEAKKKAAAEAAKKKAAVEA 204
```

RESULT 13
AAY14928
ID AAY14928 standard; protein; 223 AA.
XX
AC AAY14928;
XX
DT 25-OCT-1999 (first entry)
XX
DE Amino acid sequence of M. vaccae antigen GV-45.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI; 1999-430163/36.
DR N-PSDB; AAZ11393.
XX
PT Enhancing immune response to an antigen
XX
PS Claim 1; Page 239; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 223 AA;

Query Match 34.3%; Score 178; DB 20; Length 223;
Best Local Similarity 54.5%; Pred. No. 1.1e-07;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

```
Qy          1 AKKYAKK--AEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              |:| |||  |:|| ||||  ||:   ||| | | | | ||  ||| | |:    ||:||
Db        112 ARKAAKKAPAKKAAAKKAAPAKKAPAKKA-ATKAAPAKKATAAKKAAPAKKATAAKKAAP 170
```

```
Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAAK 103
                |||   | | | | | ||  |: |  || |||| |||  || | || ||
Db        171 AKKAPAKKAATKAAPAKKAPAKKAATKAAPAKKAPAAK--KAPAKKAPAK 218
```

RESULT 14
AAR84568
ID AAR84568 standard; Protein; 643 AA.
XX
AC AAR84568;
XX
DT 09-MAY-1996 (first entry)
XX
DE Trypanosoma cruzi TCR27 polypeptide, Ag15.
XX
KW TCR27; Chagas disease; repeat unit; diagnosis; blood screening;
KW recombinant; fusion protein; glutathione-S-transferase.
XX
OS Trypanosoma cruzi.
XX
FH Key Location/Qualifiers
FT Region 329..552
FT /label= repeat_region
FT /note= "16 of 69 repeat units of 14 amino acids"
XX
PN WO9525797-A1.
XX
PD 28-SEP-1995.
XX
PF 20-MAR-1995; 95WO-US03191.
XX
PR 24-MAR-1994; 94US-0216894.
XX
PA (KIRC/) KIRCHHOFF L V.
PA (OTSU/) OTSU K.
XX
PI Kirchhoff LV, Otsu K;
XX
DR WPI; 1995-344618/44.
DR N-PSDB; AAT05332.
XX
PT New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
PT immunoassay reagent for specific diagnosis of Chagas disease, also
PT related nucleic acid and transformed cells
XX
PS Disclosure; Page 40-41; 68pp; English.
XX
CC AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi
CC The proteins are all fusion products with glutathione-S-transferase
CC (GST) and some contain a linker sequence. The TCR27 protein comprises
CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
CC encodes the GST sequence, the Ag44 polypeptide contg. 16 of the 69
CC repeat units and also contains the amino and carboxy terminal
CC peptides of TCR27. The TCR27 polypeptides of the invention are useful
CC for the diagnosis of Chagas disease (American Trypanosomiasis), they
CC are capable of detecting anti-T.cruzi antibodies; or for blood
CC screening. The TCR27 protein has epitopes to which most T.cruzi
CC infected individuals have antibodies. The TCR27 polypeptides will not
CC react with serum from patients with leishmaniasis, schistosomiasis,
CC or autoimmune disease and are hence less likely to cause false
CC positives in diagnosis.
XX
SQ Sequence 643 AA;

Query Match 31.3%; Score 162.5; DB 16; Length 643;
Best Local Similarity 44.4%; Pred. No. 6.5e-06;
Matches 52; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

```
Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYA----KEAAKA 56
              | :  | ||    | |:| |  :|  :| |:| | |||:|:  ||| | |    ::||:|
Db        445 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 504
```

```
Qy          57 KK--EAYKAEAKKYAKAAKAEKKEYA----AAEAKKAEAAKAYK-AEAAKAAAKEAA 106
               |  || | :| :  | |:|||:: |     |||:| :||:| | ||| |  | |||
Db         505 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEAA 561
```

RESULT 15
AAY34068
ID AAY34068 standard; peptide; 158 AA.
XX
AC AAY34068;
XX
DT 23-NOV-1999 (first entry)
XX
DE Histone H1 isoform H1.5 pANCA-reactive fragment (residues 69-226).
XX
KW Ulcerative colitis; histone; H1-like antigen; porin antigen; human;
KW Bacteroides antigen; inflammatory bowel disease; IBD; pANCA; diagnosis;
KW perinuclear anti-neutrophil cytoplasmic antibody; isoform.
XX
OS Homo sapiens.
XX
PN WO9945955-A1.
XX
PD 16-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05492.
XX
PR 12-MAR-1998; 98US-0041889.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Braun J, Cohavy O;
XX
DR WPI; 1999-551215/46.
XX
PT Use of histone H1, porin or Bacteroides antigens as targets for the
PT diagnosis, prevention and treatment of ulcerative colitis -
XX
PS Example 4; Page 125-126; 134pp; English.
XX
CC The invention provides a method for the diagnosis, prevention and
CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a
CC porin antigen or a Bacteroides antigen as a target antigen. The novel
CC method of diagnosing UC in a subject suspected of having inflammatory
CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
CC (2) contacting the sample with a histone H1-like antigen, or perinuclear
CC anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
CC complex of the histone H1-like antigen, or the pANCA-reactive fragment,
CC and antibody to the histone H1-like antigen; and (3) detecting the
CC presence or absence of the complex; where the presence of the complex
CC indicates that the subject has UC. The pANCA-reactive histone H1-like
CC antigen, porin antigen and Bacteroides antigen are useful in the
CC diagnosis, prevention and treatment of UC. The methods can also be used
CC for identifying agents useful for treating UC. The present sequence
CC represents a human histone H1 isoform H1.5 pANCA-reactive fragment.
XX
SQ Sequence 158 AA;

Query Match 30.9%; Score 160.5; DB 20; Length 158;
Best Local Similarity 46.8%; Pred. No. 2.2e-06;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

```
Qy           1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAA-----EAKKKAKAEAKKYAKEAAK 55
               ||  ||||  | |||   |  ||  ||| | | ||      :||| | |  || ||   |
Db          51 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKK 108

Qy          56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
               ||  |   :| |     || | : |  :| | :|||   |:| |||||:
Db         109 AKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKK 157
```

Search completed: January 21, 2004, 08:59:10

Job time : 53.3766 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:01:02 ; Search time 39.6364 Seconds
(without alignments)
562.323 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	519	100.0	109	10	US-09-816-989A-7	Sequence 7, Appli
2	312.5	60.2	86	10	US-09-816-989A-6	Sequence 6, Appli
3	289	55.7	77	10	US-09-816-989A-5	Sequence 5, Appli
4	228.5	44.0	66	10	US-09-816-989A-4	Sequence 4, Appli
5	180.5	34.8	56	10	US-09-816-989A-3	Sequence 3, Appli
6	179	34.5	347	15	US-10-127-032-120	Sequence 120, App
7	178	34.3	223	12	US-10-205-979-52	Sequence 52, Appl
8	178	34.3	223	14	US-10-051-643-201	Sequence 201, App
9	163	31.4	376	15	US-10-156-761-9889	Sequence 9889, Ap
10	162	31.2	356	11	US-09-820-843A-27	Sequence 27, Appl
11	160.5	30.9	158	15	US-10-229-567-40	Sequence 40, Appl
12	160.5	30.9	226	15	US-10-229-567-32	Sequence 32, Appl
13	159.5	30.7	234	15	US-10-262-209-1	Sequence 1, Appli
14	158	30.4	218	15	US-10-229-567-4	Sequence 4, Appli
15	154.5	29.8	130	15	US-10-262-209-2	Sequence 2, Appli
16	154.5	29.8	309	11	US-09-820-843A-24	Sequence 24, Appl
17	153.5	29.6	369	11	US-09-820-843A-95	Sequence 95, Appl
18	153.5	29.6	372	11	US-09-820-843A-8	Sequence 8, Appli
19	152.5	29.4	222	15	US-10-229-567-3	Sequence 3, Appli
20	144	27.7	352	11	US-09-820-843A-23	Sequence 23, Appl
21	144	27.7	434	9	US-09-124-280A-3	Sequence 3, Appli
22	143.5	27.6	452	15	US-10-184-832-5	Sequence 5, Appli
23	143	27.6	214	15	US-10-229-567-27	Sequence 27, Appl
24	138.5	26.7	217	15	US-10-156-761-10221	Sequence 10221, A
25	138.5	26.7	617	9	US-09-864-761-36182	Sequence 36182, A
26	138.5	26.7	617	12	US-10-029-386-32114	Sequence 32114, A
27	138	26.6	45	10	US-09-816-989A-2	Sequence 2, Appli
28	134.5	25.9	105	12	US-10-177-725-43	Sequence 43, Appl
29	134.5	25.9	106	12	US-10-177-725-44	Sequence 44, Appl
30	134.5	25.9	524	15	US-10-128-714-3303	Sequence 3303, Ap
31	133.5	25.7	104	12	US-10-177-725-40	Sequence 40, Appl
32	133.5	25.7	212	15	US-10-229-567-1	Sequence 1, Appli
33	129.5	25.0	106	12	US-10-177-725-45	Sequence 45, Appl
34	129.5	25.0	113	15	US-10-229-567-14	Sequence 14, Appl
35	129.5	25.0	147	15	US-10-229-567-13	Sequence 13, Appl
36	129.5	25.0	220	15	US-10-229-567-2	Sequence 2, Appli
37	129	24.9	86	12	US-10-177-725-19	Sequence 19, Appl
38	128.5	24.8	104	12	US-10-177-725-41	Sequence 41, Appl
39	128	24.7	83	12	US-10-177-725-12	Sequence 12, Appl
40	128	24.7	85	12	US-10-177-725-15	Sequence 15, Appl
41	128	24.7	110	12	US-10-177-725-48	Sequence 48, Appl
42	128	24.7	188	12	US-10-259-165-278	Sequence 278, App
43	128	24.7	191	12	US-10-259-165-92	Sequence 92, Appl
44	128	24.7	191	12	US-10-259-165-372	Sequence 372, App
45	127.5	24.6	104	12	US-10-177-725-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-09-816-989A-7
; Sequence 7, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match 100.0%; Score 519; DB 10; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.9e-34;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
```

```
RESULT 2
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 86
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

  Query Match             60.2%;  Score 312.5;  DB 10;  Length 86;
  Best Local Similarity   72.1%;  Pred. No. 3.2e-18;
  Matches   80;  Conservative    1;  Mismatches    3;  Indels   27;  Gaps    5;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA------------------ 36

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAK-KAEAA-KAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||| ||||| ||||||||||||||||||
Db         37 -KAEAKKYAKAAKAEKKEYAAAEAKYKAEAAKKAYKAEAAKAAAKEAAYEA 86


RESULT 3
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 77
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

  Query Match             55.7%;  Score 289;  DB 10;  Length 77;
  Best Local Similarity   67.0%;  Pred. No. 2.1e-16;
  Matches   73;  Conservative    1;  Mismatches    3;  Indels   32;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA------------------ 36

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||        ||||||||||||||||
Db         37 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 77


RESULT 4
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

  Query Match             44.0%;  Score 228.5;  DB 10;  Length 66;
  Best Local Similarity   56.9%;  Pred. No. 1.1e-11;
  Matches   62;  Conservative    0;  Mismatches    4;  Indels   43;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||  ||          |||| | |
Db          1 AKKYAKK-EKAYAKAKKA----------EAKAAKKA----------------------- 25

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||        ||||||||||||||||
Db         26 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 66


RESULT 5
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
;  APPLICANT: Gad, Alexander
;  APPLICANT:  Lis, Doris
;  TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
;  TITLE OF INVENTION:  AND FOR THERAPEUTIC USE
;  FILE REFERENCE: 2609/60807-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/816,989A
;  CURRENT FILING DATE:  2001-03-23
;  PRIOR APPLICATION NUMBER: 60/101,693
;  PRIOR FILING DATE: 1998-09-25
;  PRIOR APPLICATION NUMBER: PCT/US99/22402
;  PRIOR FILING DATE: 1999-09-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

  Query Match             34.8%;  Score 180.5;  DB 10;  Length 56;
  Best Local Similarity   45.9%;  Pred. No. 5.8e-08;
```

```
  Matches   50;  Conservative    3;  Mismatches    3;  Indels   53;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | |||||||||||||:|:
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKAEAK-------------- 39

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
                                    ||||||||||||||||
Db         40 --------------------------------YKAEAAKAAAKEAAYEA 56


RESULT 6
US-10-127-032-120
; Sequence 120, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
;  APPLICANT: Whiteley, Marvin
;  APPLICANT:  Bangera, M. Gita
;  APPLICANT:  Lory, Stephen
;  APPLICANT:  Greenberg, Everett Peter
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
;  TITLE OF INVENTION:  BIOFILM FORMATION
;  FILE REFERENCE: UIZ-070CP
;  CURRENT APPLICATION NUMBER: US/10/127,032
;  CURRENT FILING DATE:  2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,190
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: US 60/344,142
;  PRIOR FILING DATE: 2001-10-24
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 120
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-10-127-032-120

  Query Match             34.5%;  Score 179;  DB 15;  Length 347;
  Best Local Similarity   49.1%;  Pred. No. 5e-07;
  Matches   54;  Conservative   17;  Mismatches   29;  Indels   10;  Gaps    4;

Qy          6 KKAEKAYAKKAKAAKEK------KAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              :| |:     ||||::|      || |:| |:| || |||| |:|:| :  |:|  |||
Db         99 QKLEQQQVAAAKAAEQKKADEARKAEAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKR 158

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              | : ||||  |||:  ||: || :|||  | :| |  ||:|| |:|| ||
Db        159 A-EDEAKK--KAAEDAKKK-AAEDAKKKAAAEEAKKKAAAEAAKKKAAVEA 204


RESULT 7
US-10-205-979-52
; Sequence 52, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Tan, Paul L. J.
;  APPLICANT:  Abernethy, Nevin
;  TITLE OF INVENTION: Compounds and Methods for the Modulation
;  TITLE OF INVENTION:  of Immune Responses
;  FILE REFERENCE: 11000.1063U
;  CURRENT APPLICATION NUMBER: US/10/205,979
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/308,446
;  PRIOR FILING DATE: 2001-07-26
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-10-205-979-52

  Query Match             34.3%;  Score 178;  DB 12;  Length 223;
  Best Local Similarity   54.5%;  Pred. No. 3.8e-07;
  Matches   60;  Conservative    6;  Mismatches   34;  Indels   10;  Gaps    5;

Qy          1 AKKYAKK--AEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              |:| |||  |:|| ||||  ||:   ||| | | | | ||  ||| | |:    ||:||
Db        112 ARKAAKKAPAKKAAAKKAAPAKKAPAKKA-ATKAAPAKKATAAKKAAPAKKATAAKKAAP 170

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAAK 103
              |||   |  | | | | ||  |: |  || |||| |||  || | || ||
Db        171 AKKAPAKKAATKAAPAKKAPAKKAATKAAPAKKAPAAK--KAPAKKAPAK 218


RESULT 8
US-10-051-643-201
; Sequence 201, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Tan, Paul L. J.
;  TITLE OF INVENTION: Methods and Compounds for the Treatment
;  TITLE OF INVENTION:  of Immunologically-Mediated Diseases of the Respiratory
;  TITLE OF INVENTION:  System using Mycobacterium Vaccae
;  FILE REFERENCE: 11000.1008c2
;  CURRENT APPLICATION NUMBER: US/10/051,643
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US09/156,181
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: US 08/996,624
;  PRIOR FILING DATE: 1997-12-23
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 201
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-10-051-643-201

  Query Match             34.3%;  Score 178;  DB 14;  Length 223;
  Best Local Similarity   54.5%;  Pred. No. 3.8e-07;
  Matches   60;  Conservative    6;  Mismatches   34;  Indels   10;  Gaps    5;

Qy          1 AKKYAKK--AEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              |:| |||  |:|| ||||  ||:   ||| | | | | ||  ||| | |:    ||:||
Db        112 ARKAAKKAPAKKAAAKKAAPAKKAPAKKA-ATKAAPAKKATAAKKAAPAKKATAAKKAAP 170

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAAK 103
              |||   |  | | | | ||  |: |  || |||| |||  || | || ||
Db        171 AKKAPAKKAATKAAPAKKAPAKKAATKAAPAKKAPAAK--KAPAKKAPAK 218


RESULT 9
US-10-156-761-9889
; Sequence 9889, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9889
```

```
;   LENGTH: 376
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9889

  Query Match             31.4%;  Score 163;  DB 15;  Length 376;
  Best Local Similarity   44.9%;  Pred. No. 9.9e-06;
  Matches   48;  Conservative   15;  Mismatches   38;  Indels    6;  Gaps    2;

Qy          6 KKAEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAK---KKAKAEAKKYAKEAAKAKKE 59
              : |:|| |:||  ||:   ||| ||| | | |:|  |   ||| |:    ||::|  |
Db        245 RTAKKAAARKAAPAKKTPAKKAAAKKTAPAKKSAPGKTAAKKAAAKKSAPAKKSAPGKTA 304

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
              | || ||| | | |:   : || :|   : | | |: | |:|||: |
Db        305 AKKAAAKKSAPAKKSAPGKTAAKKAAAKKTAPAKKSAAKKSAAKKTA 351


RESULT 10
US-09-820-843A-27
; Sequence 27, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
;  TITLE OF INVENTION:  USEFUL AS ANTI-INFECTIVES
;  FILE REFERENCE: Q63915
;  CURRENT APPLICATION NUMBER: US/09/820,843A
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 356
;   TYPE: PRT
;   ORGANISM: Vibrio cholerae
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: tolA protein
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: gi|9656364
US-09-820-843A-27

  Query Match             31.2%;  Score 162;  DB 11;  Length 356;
  Best Local Similarity   42.6%;  Pred. No. 1.1e-05;
  Matches   55;  Conservative   13;  Mismatches   37;  Indels   24;  Gaps    5;

Qy          3 KYAKKAAEKAYAKK--AKAAKEKKAYAKKEAKAYKAAEAKKK----AKAEAKKYAKEAAKA 56
              | |::|||   :|   : | |:||  :||  |   || | |      ||| :: ||||| |
Db        112 KAAREAAKLREQKEQERLAAEQKAREEKERAAKAEAERKVKEEAAKKAEQERVAKEAAAA 171

Qy         57 KKEAYKAEAKKYAK------------AAKAEK----KEYAAAEAKKAEAAKAYKAEAAKA 100
              | |  : | :| ||            |||||:    || || ||  |: ||  |  ||||
Db        172 KAEQQRIEREKEAKLAEEKAKREKEVAAKAEQERLAKEKAAKEA--ADKAKKEKERAAKA 229

Qy        101 AAKEAAYEA 109
               |:  | ||
Db        230 EAERKAQEA 238


RESULT 11
US-10-229-567-40
; Sequence 40, Application US/10229567
; Publication No. US20030092080A1
;   GENERAL INFORMATION:
;        APPLICANT: Braun, Jonathan
;                   Cohavy, Offer
;        TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;                            Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;                            Microbial UC pANCA antigens
;        NUMBER OF SEQUENCES: 41
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Campbell & Flores LLP
;             STREET: 4370 La Jolla Village Drive, Suite 700
;             CITY: San Diego
;             STATE: California
;             COUNTRY: USA
;             ZIP: 92122
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/229,567
;             FILING DATE: 27-Aug-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/417,264
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: US 09/041,889
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Campbell, Cathryn A.
;             REGISTRATION NUMBER: 31,815
;             REFERENCE/DOCKET NUMBER: P-PM 3006
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (619) 535-9001
;             TELEFAX: (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 40:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 158 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-229-567-40

  Query Match             30.9%;  Score 160.5;  DB 15;  Length 158;
  Best Local Similarity   46.8%;  Pred. No. 6.4e-06;
  Matches   51;  Conservative    7;  Mismatches   44;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAA-----EAKKKAKAEAKKYAKEAAK 55
              ||  |||| | |||    |  ||  ||| | | ||      :||| | |  || ||   |
Db         51 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKK 108

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              ||  |   :| |      || | : |  :| | :|||   |:| |||||:
Db        109 AKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKK 157


RESULT 12
US-10-229-567-32
; Sequence 32, Application US/10229567
; Publication No. US20030092080A1
;   GENERAL INFORMATION:
;        APPLICANT: Braun, Jonathan
;                   Cohavy, Offer
;        TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;                            Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;                            Microbial UC pANCA antigens
;        NUMBER OF SEQUENCES: 41
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Campbell & Flores LLP
;             STREET: 4370 La Jolla Village Drive, Suite 700
;             CITY: San Diego
;             STATE: California
;             COUNTRY: USA
;             ZIP: 92122
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/229,567
;             FILING DATE: 27-Aug-2002
```

```
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/417,264
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: US 09/041,889
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Campbell, Cathryn A.
;            REGISTRATION NUMBER: 31,815
;            REFERENCE/DOCKET NUMBER: P-PM 3006
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (619) 535-9001
;            TELEFAX: (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 32:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 226 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-229-567-32

  Query Match             30.9%;  Score 160.5;  DB 15;  Length 226;
  Best Local Similarity   46.8%;  Pred. No. 9.3e-06;
  Matches   51;  Conservative    7;  Mismatches   44;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAA-----EAKKKAKAEAKKYAKEAAK 55
              ||  |||| | |||   |  ||  ||| | | ||      :||| | |  || ||   |
Db        119 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKK 176

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              ||  |   :| |     || | : |  :| | :|||   |:| |||||:
Db        177 AKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKK 225


RESULT 13
US-10-262-209-1
; Sequence 1, Application US/10262209
; Publication No. US20030125239A1
; GENERAL INFORMATION:
;  APPLICANT: Crisanti, Andrea
;  APPLICANT:  Esseghir, Selma
;  TITLE OF INVENTION: Compositions for Drug Delivery
;  FILE REFERENCE: GJE-6703
;  CURRENT APPLICATION NUMBER: US/10/262,209
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: UK 0218324.2
;  PRIOR FILING DATE: 2002-08-07
;  PRIOR APPLICATION NUMBER: PCT/GB01/01699
;  PRIOR FILING DATE: 2001-04-12
;  PRIOR APPLICATION NUMBER: UK 0102667.3
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: UK 0009080.3
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 234
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-262-209-1

  Query Match             30.7%;  Score 159.5;  DB 15;  Length 234;
  Best Local Similarity   45.7%;  Pred. No. 1.2e-05;
  Matches   53;  Conservative    8;  Mismatches   44;  Indels   11;  Gaps    3;

Qy          2 KKYAKKAEKAYAKKAKAAKEKK--AYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              || |    |  |||| ||| ||    |||  ||  ||  || ||   ||  | || |
Db        109 KKAASGEAKPKAKKAGAAKAKKPAGAAKKPKKATGAATPKKSAKKTPKKAKKPAAAA--G 166

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAK-------KAEAAKAYKAEAAKAAAKEAAYE 108
              | ||:: | ||||| :|   : |:||       | : ||   |:  |||||:   |
Db        167 AKKAKSPKKAKAAKPKKAPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAAAKKKKLE 222


RESULT 14
US-10-229-567-4
; Sequence 4, Application US/10229567
; Publication No. US20030092080A1
;  GENERAL INFORMATION:
;       APPLICANT: Braun, Jonathan
;                  Cohavy, Offer
;       TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;                           Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;                           Microbial UC pANCA antigens
;       NUMBER OF SEQUENCES: 41
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Campbell & Flores LLP
;            STREET: 4370 La Jolla Village Drive, Suite 700
;            CITY: San Diego
;            STATE: California
;            COUNTRY: USA
;            ZIP: 92122
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/229,567
;            FILING DATE: 27-Aug-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/417,264
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: US 09/041,889
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Campbell, Cathryn A.
;            REGISTRATION NUMBER: 31,815
;            REFERENCE/DOCKET NUMBER: P-PM 3006
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (619) 535-9001
;            TELEFAX: (619) 535-8949
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 218 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       FEATURE:
;            NAME/KEY:  Peptide
;            LOCATION:  1..218
;            OTHER INFORMATION:  /note= "product = Human Histone
;            H1-S-4"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-229-567-4

  Query Match             30.4%;  Score 158;  DB 15;  Length 218;
  Best Local Similarity   48.6%;  Pred. No. 1.4e-05;
  Matches   53;  Conservative    7;  Mismatches   43;  Indels    6;  Gaps    3;

Qy          2 KKYAKKAEKAYAKKAKAAKEKK--AYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              || |    |  |||| ||| ||    |||  ||  ||  || ||   ||  | || |
Db        108 KKAASGEAKPKAKKAGAAKAKKPAGAAKKPKKATGAATPKKSAKKTPKKAKKPAAAA--G 165

Qy         60 AYKAEAKKYAKAAKAEK--KEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
              | ||:: | ||||| :|   |  | |:| | :|||   |:   |  |:||
Db        166 AKKAKSPKKAKAAKPKKAPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAA 214


RESULT 15
US-10-262-209-2
; Sequence 2, Application US/10262209
; Publication No. US20030125239A1
; GENERAL INFORMATION:
;  APPLICANT: Crisanti, Andrea
```

```
;  APPLICANT:  Esseghir, Selma
;  TITLE OF INVENTION: Compositions for Drug Delivery
;  FILE REFERENCE: GJE-6703
;  CURRENT APPLICATION NUMBER: US/10/262,209
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: UK 0218324.2
;  PRIOR FILING DATE: 2002-08-07
;  PRIOR APPLICATION NUMBER: PCT/GB01/01699
;  PRIOR FILING DATE: 2001-04-12
;  PRIOR APPLICATION NUMBER: UK 0102667.3
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: UK 0009080.3
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 130
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-262-209-2

  Query Match             29.8%;  Score 154.5;  DB 15;  Length 130;
  Best Local Similarity   47.3%;  Pred. No. 1.6e-05;
  Matches   52;  Conservative    8;  Mismatches   43;  Indels    7;  Gaps    3;

Qy          2 KKYAKKAEKAYAKK---AKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKK 58
              || |    |  |||   |||||:    |||  ||  ||  || ||   ||  | || |
Db         17 KKAASGEAKPKAKKAGAAKAAKKPAGAAKKPKKATGAATPKKSAKKTPKKAKKPAAAA-- 74

Qy         59 EAYKAEAKKYAKAAKAEK--KEYAAAEAKKAEAAKAYKAEAAKAAAKEAA 106
                | ||:: | ||||| :|  |  | |:| | :|||   |:   |  |:||
Db         75 GAKKAKSPKKAKAAKPKKAPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAA 124


Search completed: January 21, 2004, 09:10:08
Job time : 39.6364 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:57:47 ; Search time 20.526 Seconds
(without alignments)
224.685 Million cell updates/sec

Title: US-09-816-989A-7
Perfect score: 519
Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAAKEAAYEA 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	519	100.0	109	4	US-09-405-743A-7	Sequence 7, Appli
2	312.5	60.2	86	4	US-09-405-743A-6	Sequence 6, Appli
3	289	55.7	77	4	US-09-405-743A-5	Sequence 5, Appli
4	228.5	44.0	66	4	US-09-405-743A-4	Sequence 4, Appli
5	180.5	34.8	56	4	US-09-405-743A-3	Sequence 3, Appli
6	179.5	34.6	100	2	US-08-460-890A-64	Sequence 64, Appl
7	179.5	34.6	100	3	US-08-167-641C-64	Sequence 64, Appl
8	179.5	34.6	100	3	US-08-460-971A-64	Sequence 64, Appl
9	179.5	34.6	100	3	US-08-462-040-64	Sequence 64, Appl
10	179	34.5	407	4	US-09-252-991A-29581	Sequence 29581, A
11	178	34.3	223	3	US-09-095-855-201	Sequence 201, App
12	178	34.3	223	4	US-09-205-426-201	Sequence 201, App
13	162.5	31.3	643	2	US-08-216-894-8	Sequence 8, Appli
14	162.5	31.3	643	3	US-09-115-746-8	Sequence 8, Appli
15	160.5	30.9	158	3	US-09-041-889-40	Sequence 40, Appl
16	160.5	30.9	158	4	US-09-417-264-40	Sequence 40, Appl
17	160.5	30.9	226	3	US-09-041-889-32	Sequence 32, Appl
18	160.5	30.9	226	4	US-09-417-264-32	Sequence 32, Appl
19	160.5	30.9	468	4	US-09-328-352-6321	Sequence 6321, Ap
20	160	30.8	214	4	US-09-328-352-5169	Sequence 5169, Ap
21	158.5	30.5	472	2	US-08-216-894-10	Sequence 10, Appl
22	158.5	30.5	472	3	US-09-115-746-10	Sequence 10, Appl
23	158.5	30.5	564	2	US-08-216-894-2	Sequence 2, Appli
24	158.5	30.5	564	3	US-09-115-746-2	Sequence 2, Appli
25	158	30.4	218	3	US-09-041-889-4	Sequence 4, Appli
26	158	30.4	218	3	US-08-837-058-4	Sequence 4, Appli
27	158	30.4	218	4	US-09-417-264-4	Sequence 4, Appli
28	152.5	29.4	222	3	US-09-041-889-3	Sequence 3, Appli
29	152.5	29.4	222	3	US-08-837-058-3	Sequence 3, Appli
30	152.5	29.4	222	4	US-09-417-264-3	Sequence 3, Appli
31	151	29.1	316	4	US-09-252-991A-32957	Sequence 32957, A
32	150	28.9	1507	3	US-08-929-329-5	Sequence 5, Appli
33	144	27.7	399	4	US-09-252-991A-22853	Sequence 22853, A
34	144	27.7	434	1	US-08-097-830E-3	Sequence 3, Appli
35	144	27.7	434	2	US-08-456-112B-3	Sequence 3, Appli
36	143	27.6	214	3	US-09-041-889-27	Sequence 27, Appl
37	143	27.6	214	4	US-09-417-264-27	Sequence 27, Appl
38	141	27.2	100	2	US-08-460-890A-62	Sequence 62, Appl
39	141	27.2	100	3	US-08-167-641C-62	Sequence 62, Appl
40	141	27.2	100	3	US-08-460-971A-62	Sequence 62, Appl
41	141	27.2	100	3	US-08-462-040-62	Sequence 62, Appl
42	138	26.6	45	4	US-09-405-743A-2	Sequence 2, Appli
43	133.5	25.7	212	3	US-09-041-889-1	Sequence 1, Appli
44	133.5	25.7	212	3	US-08-837-058-1	Sequence 1, Appli
45	133.5	25.7	212	4	US-09-417-264-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-405-743A-7
; Sequence 7, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A
; CURRENT APPLICATION NUMBER: US/09/405,743A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-405-743A-7

Query Match 100.0%; Score 519; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.7e-37;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
```

RESULT 2
US-09-405-743A-6
; Sequence 6, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A
; CURRENT APPLICATION NUMBER: US/09/405,743A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence

```
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-6

  Query Match             60.2%;  Score 312.5;  DB 4;  Length 86;
  Best Local Similarity   72.1%;  Pred. No. 6.5e-20;
  Matches   80;  Conservative    1;  Mismatches    3;  Indels   27;  Gaps    5;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA------------------ 36

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAK-KAEAA-KAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||| ||||| ||||||||||||||||||
Db         37 -KAEAKKYAKAAKAEKKEYAAAEAKYKAEAAKKAYKAEAAKAAAKEAAYEA 86


RESULT 3
US-09-405-743A-5
; Sequence 5, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 77
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-5

  Query Match             55.7%;  Score 289;  DB 4;  Length 77;
  Best Local Similarity   67.0%;  Pred. No. 5.4e-18;
  Matches   73;  Conservative    1;  Mismatches    3;  Indels   32;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKA------------------ 36

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||       ||||||||||||||||||
Db         37 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 77


RESULT 4
US-09-405-743A-4
; Sequence 4, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 66
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-4

  Query Match             44.0%;  Score 228.5;  DB 4;  Length 66;
  Best Local Similarity   56.9%;  Pred. No. 5.2e-13;
  Matches   62;  Conservative    0;  Mismatches    4;  Indels   43;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||      ||          |||| | |
Db          1 AKKYAKK-EKAYAKKA----------EAKAAKKA------------------------- 25

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              ||||||||||||||||||||||||       ||||||||||||||||||
Db         26 -KAEAKKYAKAAKAEKKEYAAAEAK-------YKAEAAKAAAKEAAYEA 66


RESULT 5
US-09-405-743A-3
; Sequence 3, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-3

  Query Match             34.8%;  Score 180.5;  DB 4;  Length 56;
  Best Local Similarity   45.9%;  Pred. No. 4.5e-09;
  Matches   50;  Conservative    3;  Mismatches    3;  Indels   53;  Gaps    3;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEA 60
              ||||||| ||||||||     :||  | ||||||||||||||:|:
Db          1 AKKYAKK-EKAYAKKA-----EKAAKKAEAKAYKAAEAKKKAEAK-------------- 39

Qy         61 YKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
                                                ||||||||||||||||
Db         40 --------------------------------YKAEAAKAAAKEAAYEA 56


RESULT 6
US-08-460-890A-64
; Sequence 64, Application US/08460890A
; Patent No. 5994109
;  GENERAL INFORMATION:
;    APPLICANT:  Woo, Savio L.C.
;    APPLICANT:  Smith, Louis C.
;    APPLICANT:  Cristiano, Richard J.
;    APPLICANT:  Gottchalk, Stephen
;    TITLE OF INVENTION:  NUCLEIC ACID TRANSPORTER SYSTEMS AND
;    TITLE OF INVENTION:  METHODS OF USE
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,890A
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167,641
;      FILING DATE:  December 14, 1993
;      APPLICATION NUMBER:  07/855,389
;      FILING DATE:  March 20, 1992
;      APPLICATION NUMBER:  PCT/US93/02725
;      FILING DATE:  March 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/066
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  100 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  "Lys Ala" in positions 3 to 100 may be
;      OTHER INFORMATION:  present or absent.
US-08-460-890A-64

  Query Match             34.6%;  Score 179.5;  DB 2;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 9.7e-09;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 7
US-08-167-641C-64
; Sequence 64, Application US/08167641C
; Patent No. 6033884
;  GENERAL INFORMATION:
;    APPLICANT:  Woo, Savio L.C.
;    APPLICANT:  Smith, Louis C.
;    APPLICANT:  Cristiano, Richard J.
;    APPLICANT:  Gottchalk, Stephen
;    TITLE OF INVENTION:  NUCLEIC ACID TRANSPORTER SYSTEMS AND
;    TITLE OF INVENTION:  METHODS OF USE
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/167,641C
;      FILING DATE:  December 14, 1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/855,389
;      FILING DATE:  March 20, 1992
;      APPLICATION NUMBER:  PCT/US93/02725
;      FILING DATE:  March 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  205/012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  100 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  "Lys Ala" in positions 3 to 100 may be
;      OTHER INFORMATION:  present or absent.
US-08-167-641C-64

  Query Match             34.6%;  Score 179.5;  DB 3;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 9.7e-09;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 8
US-08-460-971A-64
; Sequence 64, Application US/08460971A
; Patent No. 6150168
;  GENERAL INFORMATION:
;    APPLICANT:  Woo, Savio L.C.
;    APPLICANT:  Smith, Louis C.
;    APPLICANT:  Cristiano, Richard J.
;    APPLICANT:  Gottchalk, Stephen
;    TITLE OF INVENTION:  NUCLEIC ACID TRANSPORTER SYSTEMS AND
;    TITLE OF INVENTION:  METHODS OF USE
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,971A
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167,641
;      FILING DATE:  December 14, 1993
;      APPLICATION NUMBER:  07/855,389
```

```
;       FILING DATE:  March 20, 1992
;       APPLICATION NUMBER:  PCT/US93/02725
;       FILING DATE:  March 19, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  212/063
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  100 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  "Lys Ala" in positions 3 to 100 may be
;       OTHER INFORMATION:  present or absent.
US-08-460-971A-64

  Query Match             34.6%;  Score 179.5;  DB 3;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 9.7e-09;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 9
US-08-462-040-64
; Sequence 64, Application US/08462040
; Patent No. 6177554
;  GENERAL INFORMATION:
;    APPLICANT:  Woo, Savio L.C.
;    APPLICANT:  Smith, Louis C.
;    APPLICANT:  Cristiano, Richard J.
;    APPLICANT:  Gottchalk, Stephen
;    TITLE OF INVENTION:  NUCLEIC ACID TRANSPORTER SYSTEMS AND
;    TITLE OF INVENTION:  METHODS OF USE
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,040
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167,641
;      FILING DATE:  December 14, 1993
;      APPLICATION NUMBER:  07/855,389
;      FILING DATE:  March 20, 1992
;      APPLICATION NUMBER:  PCT/US93/02725
;      FILING DATE:  March 19, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  100 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      OTHER INFORMATION:  "Lys Ala" in positions 3 to 100 may be
;      OTHER INFORMATION:  present or absent.
US-08-462-040-64

  Query Match             34.6%;  Score 179.5;  DB 3;  Length 100;
  Best Local Similarity   55.0%;  Pred. No. 9.7e-09;
  Matches   55;  Conservative    8;  Mismatches   34;  Indels    3;  Gaps    3;

Qy          6 KKAEKAYAK-KAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKEAYKAE 64
              |   || || ||||  : || || :|||   |:|| ||||:||  ||  |||| :| ||:
Db          1 KAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59

Qy         65 AKKYAKA-AKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAK 103
              ||  ||| |||: |  | |:||    |||     ||| ||
Db         60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99


RESULT 10
US-09-252-991A-29581
; Sequence 29581, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29581
;   LENGTH: 407
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29581

  Query Match             34.5%;  Score 179;  DB 4;  Length 407;
  Best Local Similarity   49.1%;  Pred. No. 4.4e-08;
  Matches   54;  Conservative   17;  Mismatches   29;  Indels   10;  Gaps    4;

Qy          6 KKAEKAYAKKAKAAKEK------KAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAKAKKE 59
              :| |:     ||||::|      || |:| |:| || |||| |:|:| :  |:|  |||
Db        159 QKLEQQQVAAAKAAEQKKADEARKAEAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKR 218

Qy         60 AYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKEAAYEA 109
              | : ||||  |||:  ||: || :|||  | :| |  ||:|| |:|| ||
Db        219 A-EDEAKK--KAAEDAKKK-AAEDAKKKAAEEAKKKAAAEAAKKKAAVEA 264


RESULT 11
US-09-095-855-201
; Sequence 201, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  Compounds and Methods for
;    TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/095,855
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/705,347
;      FILING DATE:  29-AUG-1996
;      APPLICATION NUMBER:  08/873,970
;      FILING DATE:  12-JUN-1997
;      APPLICATION NUMBER:  08/997,362
;      FILING DATE:  23-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  223 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-095-855-201

  Query Match             34.3%;  Score 178;  DB 3;  Length 223;
  Best Local Similarity   54.5%;  Pred. No. 2.9e-08;
  Matches   60;  Conservative    6;  Mismatches   34;  Indels   10;  Gaps    5;

Qy          1 AKKYAKK--AEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              |:| |||  |:|| ||||  ||:   ||| | | | | ||  ||| | |:    ||:||
Db        112 ARKAAKKAPAKKAAAKKAAPAKKAPAKKA-ATKAAPAKKATAAKKAAPAKKATAAKKAAP 170

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAAK 103
                |||   | | | | ||  |: |  || |||| |||  || | || ||
Db        171 AKKAPAKKAATKAAPAKKAPAKKAATKAAPAKKAPAAK--KAPAKKAPAK 218


RESULT 12
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Tan, Paul L. J.
;  TITLE OF INVENTION: Compounds and Methods for Treatment and
;  TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
;  FILE REFERENCE: 11000.1002c4
;  CURRENT APPLICATION NUMBER: US/09/205,426
;  CURRENT FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: 09/095,855
;  EARLIER FILING DATE: 1998-06-11
;  EARLIER APPLICATION NUMBER: 08/997,362
;  EARLIER FILING DATE: 1997-12-23
;  EARLIER APPLICATION NUMBER: 08/873,970
;  EARLIER FILING DATE: 1997-06-12
;  EARLIER APPLICATION NUMBER: 08/705,347
;  EARLIER FILING DATE: 1996-08-29
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 201
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Mycobacterium vaccae
US-09-205-426-201

  Query Match             34.3%;  Score 178;  DB 4;  Length 223;
  Best Local Similarity   54.5%;  Pred. No. 2.9e-08;
  Matches   60;  Conservative    6;  Mismatches   34;  Indels   10;  Gaps    5;

Qy          1 AKKYAKK--AEKAYAKKAKAAKE---KKAYAKKEAKAYKAAEAKKKAKAEAKKYAKEAAK 55
              |:| |||  |:|| ||||  ||:   ||| | | | | ||  ||| | |:    ||:||
Db        112 ARKAAKKAPAKKAAAKKAAPAKKAPAKKA-ATKAAPAKKATAAKKAAPAKKATAAKKAAP 170

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAAK 103
                |||   | | | | ||  |: |  || |||| |||  || | || ||
Db        171 AKKAPAKKAATKAAPAKKAPAKKAATKAAPAKKAPAAK--KAPAKKAPAK 218


RESULT 13
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
;    APPLICANT:  Kirchhoff, Louis V.
;    APPLICANT:  Otsu, Keiko
;    TITLE OF INVENTION:  POLYPEPTIDES FOR DIAGNOSING INFECTION
;    TITLE OF INVENTION:  WITH TRYPANOSOMA CRUZI
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/216,894
;      FILING DATE:  24-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  85326/102/DRLO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  643 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-216-894-8

  Query Match             31.3%;  Score 162.5;  DB 2;  Length 643;
  Best Local Similarity   44.4%;  Pred. No. 1.6e-06;
  Matches   52;  Conservative   18;  Mismatches   36;  Indels   11;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYA----KEAAKA 56
```

```
                | :  | ||    | |:| |  :|  :| |:| | |||:|:  ||| | |    ::||:|
Db        445 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 504

Qy         57 KK--EAYKAEAKKYAKAAKAEKKEYA----AAEAKKAEAAKAYK-AEAAKAAAKEAA 106
               |  || | :| :  | |:|||:: |     |||:| :||:| | ||| |  | |||
Db        505 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEAA 561


RESULT 14
US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
;  GENERAL INFORMATION:
;    APPLICANT:  Kirchhoff, Louis V.
;    APPLICANT:  Otsu, Keiko
;    TITLE OF INVENTION:  POLYPEPTIDES FOR DIAGNOSING INFECTION
;    TITLE OF INVENTION:  WITH TRYPANOSOMA CRUZI
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/115,746
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/216,894
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  85326/102/DRLO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  643 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-115-746-8

  Query Match             31.3%;  Score 162.5;  DB 3;  Length 643;
  Best Local Similarity   44.4%;  Pred. No. 1.6e-06;
  Matches   52;  Conservative   18;  Mismatches   36;  Indels   11;  Gaps    4;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAAEAKKKAKAEAKKYA----KEAAKA 56
              | :  | ||    | |:| |  :|  :| |:| | |||:|:  ||| | |    ::||:|
Db        445 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 504

Qy         57 KK--EAYKAEAKKYAKAAKAEKKEYA----AAEAKKAEAAKAYK-AEAAKAAAKEAA 106
               |  || | :| :  | |:|||:: |     |||:| :||:| | ||| |  | |||
Db        505 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEAA 561


RESULT 15
US-09-041-889-40
; Sequence 40, Application US/09041889
; Patent No. 6033864
;  GENERAL INFORMATION:
;    APPLICANT:  Braun, Jonathan
;    APPLICANT:  Cohavy, Offer
;    TITLE OF INVENTION:  Diagnosis, Prevention and Treatment of
;    TITLE OF INVENTION:  Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;    TITLE OF INVENTION:  Microbial UC pANCA antigens
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,889
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/837,058
;      FILING DATE:  11-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-PM 3006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  158 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-041-889-40

  Query Match             30.9%;  Score 160.5;  DB 3;  Length 158;
  Best Local Similarity   46.8%;  Pred. No. 5.9e-07;
  Matches   51;  Conservative    7;  Mismatches   44;  Indels    7;  Gaps    2;

Qy          1 AKKYAKKAEKAYAKKAKAAKEKKAYAKKEAKAYKAA-----EAKKKAKAEAKKYAKEAAK 55
              ||  |||| | |||    |  ||  ||| | | ||       :||| | |  || ||    |
Db         51 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKK 108

Qy         56 AKKEAYKAEAKKYAKAAKAEKKEYAAAEAKKAEAAKAYKAEAAKAAAKE 104
              ||  |   :| |      || | : |  :| | :|||   |:| |||||:
Db        109 AKAAAKPKKATKSPAKPKAVKPKAAKPKAAKPKAAKPKAAKAKKAAAKK 157


Search completed: January 21, 2004, 09:02:21
Job time : 21.526 secs
```